Towards CNN-Based Registration of Craniocaudal and Mediolateral Oblique 2-D X-ray Mammographic Images

William C. Walton, Seung-Jun Kim, Susan C. Harvey, Lisa A. Mullen, David W. Porter

Abstract— **We investigate methodologies for the automated registration of pairs of 2-D X-ray mammographic images, taken from the two standard mammographic angles. We present two exploratory techniques, based on Convolutional Neural Networks, to examine their potential for co-registration of findings on the two standard mammographic views. To test algorithm performance, our analysis uses a synthetic, surrogate data set for performing controlled experiments, as well as real 2-D X-ray mammogram imagery. The preliminary results are promising, and provide insights into how the proposed techniques may support multi-view X-ray mammography image registration currently and as technology evolves in the future.**

I. Introduction

We consider the area of standard two-view X-ray mammography and explore the use of two Convolutional Neural Network (CNN) based approaches for performing automated registration of features from the Craniocaudal (CC) and Mediolateral Oblique (MLO) standard mammographic views.

Breast cancer is one of the leading causes of death for women worldwide with a half million lives lost annually, including 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 will be invaluable.

X-ray based two-view mammography is the main imaging modality used for breast cancer screening in asymptomatic women and is also used for more specialized diagnostic exams, which are performed when suspicious findings or symptoms are present [3]. Conventional mammography involves two-dimensional Full-Field Digital Mammography (FFDM) or, in recent years, Digital Breast Tomosynthesis (DBT). DBT is a new type of digital mammography which was FDA approved in the United States in 2011. This technique involves obtaining numerous mammographic images across an arc. Reconstruction generates multiple contiguous 1 mm thick slices through the breast, as well as synthesized 2-D images of the entire breast [4]. DBT images, like FFDM images, are obtained in standard CC and MLO views. Other modalities such as Ultrasound (US), Magnetic Resonance Imaging (MRI), Positron Emission Mammography (PEM) and Molecular Breast Imaging (MBI) can also be used to image the breast, but X-ray based mammography is the only imaging modality that has been proven to improve outcomes and decrease mortality rates when used as a screening tool.

Mammographic imaging typically involves imaging the breast from at least two different angles. The most frequently used views are the CC and MLO views. The name of each view describes the direction of the X-ray beam from the source through the breast to the X-ray detector. Thus, the CC view is obtained at an angle of 0 degrees 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]. Each view involves physically positioning and compressing the breast between two compression plates immediately adjacent to an X-ray source and detector. The purpose of the two views is to include as much breast tissue as possible, and also to locate lesions by triangulating from these projections. Breast lesions may be visible in both views or only on one view depending on the lesion location in the breast and also depending on the density of the breast tissue. When breast tissue is very dense, meaning 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, in essence hiding the finding. This is in contrast to mainly fatty breast tissue where lesions have much greater density compared to the fatty tissue, based on the attenuation of the X-ray beam as it travels through breast tissue, making the lesions readily visible.

We hypothesize that image registration will be critical in improving clinicians' accuracy and machine learning algorithms by assisting with lesion location on CC and MLO mammographic views. Currently radiologists analyze images by extrapolating between the two views in search of abnormalities. Seeing a lesion on both views is an important feature which signals to the radiologist that the lesion is more likely to be real rather than a false alarm. Additionally, in order to better characterize breast lesions, visualizing the finding in two views is beneficial. Finally, identifying a lesion in both views localizes the finding in the breast, which is

W. C. Walton is a Ph.D. student in the Computer Science and Electrical Engineering Department at the University of Maryland – Baltimore County, Maryland, USA. He is also a member of the professional staff at the Johns Hopkins University Applied Physics Laboratory (JHU/APL), Laurel MD. USA. (e-mail: wwalton1@umbc.edu).

S.-J. Kim is an Assistant Professor in the Computer Science and Electrical Engineering Department at the University of Maryland – Baltimore County, Maryland, USA. (e-mail: sjkim@umbc.edu).

S. C. Harvey is the Vice President of Medical Affairs at Hologic in the division of Breast and Skeletal Health. (e-mail: susan.harvey@hologic.com). She is formerly the Director of Johns Hopkins Breast Imaging at Johns Hopkins Medicine, in Baltimore, MD. USA and was in this role during this work. (e-mail: sharvey7@jhmi.edu).

L.A. Mullen is an Assistant Professor of Radiology in the Breast Imaging Division at Johns Hopkins Medicine, Baltimore, MD. USA. (e-mail: lmullen1@jhmi.edu).

D. W. Porter is a member of the professional staff at the Johns Hopkins University Applied Physics Laboratory (JHU/APL), Laurel MD. USA. (e-mail: david.w.porter@jhuapl.edu).

This work was supported in part by NSF grant 1631838.

critical. Thus, precise registration assists clinicians in locating findings, confirming accurate lesion detection, and therefore planning further breast imaging evaluation. Registration is essential to guide biopsies and surgical procedures as accurate information regarding lesion position is required. Machine learning algorithms and Computer Aided Diagnosis (CAD) processes that involve joint processing (or fusion) of breast images, such as is reported in [6], would also be improved by accurate registration of breast lesions.

However, the automated registration of mammographic images has proven to be a challenging task due to the non-rigid heterogeneous nature of breast tissue and due to tissue distortion that can occur as part of breast imaging, including mammographic compression [7]. Moreover, the resulting pixel-wise mappings may not be bijective, but rather involve one-to-many pixel mappings for each pixel [8]. While advancements in deep learning have generally resulted in numerous improvements in medical image processing, recent surveys indicate that a best approach has not yet been identified for medical image registration and that challenges remain in achieving the desired levels of accuracy [9] [10]. Moreover, not many automated registration methods of X-ray mammographic images have been reported to the best of our knowledge.

In this work, we explore two CNN-based techniques which may be useful for registering the CC and MLO X-ray mammographic images. One methodology evaluates pixel level registration using a CNN-based nonrigid deformation field registration approach. The second involves object level feature correspondence based on the fusion of dual Region-based CNNs [11].

This paper is organized as follows. Section II provides a brief review of existing automated medical and breast image registration techniques along with recent image registration efforts involving CNNs. Section III discusses our approaches and methodologies. In Section IV, we discuss experimental results, and Section V presents conclusions and future work.

II. Review of Existing Registration Techniques

A. Surveys and Registration Basics for Medical Imagery

There are several surveys on medical image registration [12] [13] [14], but little has been presented related to breast image registration, especially for X-ray imaging. One prominent survey by Guo et al. is devoted exclusively to breast image registration [7]. Though it predates recent developments of deep learning, it provides significant insight into key techniques and challenges associated with breast image registration. Further, it briefly discusses X-ray breast image registration, which is the focus of our study.

Guo et al. defines the problem of registration as one of finding an optimal transformation or mapping for aligning one image to another (or relating the points in one image to corresponding points in another). Similar to other registration surveys [12] [15], they describe the key components of registration as a transformation, an optimization technique, a similarity measure, and a feature space. In general, the transformation can be rigid, non-rigid, or hybrid. Various optimization techniques exist such as gradient descent-based techniques. Common similarity measures include the Sum Squared Error (SSE), Mean Absolute Error (MAE), Normalized Cross Correlation (NCC), Mutual Information (MI), and Joint Entropy [16], where the latter two are particularly suggested for intermodal registration. Depending on the feature space, the techniques can be categorized into intensity versus feature-based techniques, with features comprising either control points or various structures of the breast. The survey also notes the use of biomechanical models that may involve displacement fields or Finite Element Models (FEM) for registering breasts.

B. Non-CNN Based Breast Registration Techniques

Prior to the wide-spread use of CNNs, some examples of automated breast image registration techniques involved registering 2-D X-ray mammogram images to MRI [17] [18]. Others involved utilizing 3-D information from MRI or FEMs to aid in the registration of different 2-D X-ray mammography images [8] [19]. Several efforts proposed techniques for identifying location correspondence between CC and MLO views [20] [21] [22].

C. CNN and Deep Learning Based Registration Techniques

While a variety of research on CNN based medical image registration has been reported [23] [24] [25] [26] [27] [28] [29], our review of the literature revealed a limited amount of information regarding mammography registration. One study proposed a dual CNN architecture-based technique, similar to one of our approaches, in which two CNNs are used in parallel, one for processing each image view, whose outputs are fused [30]. A key distinction from the technique in [30] is that ours involves a different type of parallel CNN architecture and also utilizes a third CNN, versus a fully connected network, for combining the results from the individual view CNN paths.

One of the common techniques used by many CNN-based medical registration algorithms, involves non-rigid, deformation field-based approaches [23] [25] [27]. This is the basis for the first of our two proposed methods. Some of these approaches are parametric, involving the use of B-splines, or other spline techniques, while others involve the non-parametric generation of deformation fields. In general, these involve optimizing a cost function that includes an image similarity metric S for two images $I_1(\boldsymbol{x})$ and $I_2(\boldsymbol{x})$, where one of the images is warped through a deformation field $D(\boldsymbol{x})$, and a regularizer R with a nonnegative weight λ as in

$$\min_{\boldsymbol{D}} S\{I_1(\boldsymbol{x}), I_2\big(D(\boldsymbol{x})\big)\} + \lambda R(D) \quad (1)$$

The regularizer component helps to ensure that the resulting deformation field is practical [31] [32]. The regularizer is normally based on the L_2 or L_1 norm of the gradients of the deformation field, where L_1 based techniques, known as Total Variation Regularization (TVR) are more suitable for handling large, non-smooth displacements [33].

A key advantage of deformation field-based registration approaches is that they can more practically address the non-rigid, non-linear distortions that are common to medical imagery. We leverage a technique similar to that reported in [27] for one of our approaches as discussed in the next section.

III. Methodologies

For our research, we explore developments in two areas of CNNs: one involving CNN non-rigid registration algorithms

based on non-parametric deformation field generation for pixel level registration and another technique, involving the joint processing of region-based CNNs for facilitating the feature correspondence between objects in the CC and MLO views.

A. *Pixel-Level Registration*

Our approach for pixel-based registration is based on an adaptation of a Fully Convolutional Network (FCN) [34] [27]. Our version of the network, shown in Figure 1, involves a skip architecture [34]. Note that in our diagram, convolution layers also involve rectified linear units (ReLus) and batch normalization (except for the final convolution layers along each path, for which the feature maps serve as optimized deformation field components).

The network input involves two channels: the image to be registered and the target image. While this may appear to be unusual, it becomes clearer when we consider that the objective of the network is to learn a mapping, or a deformation field between the two images, as opposed to learning the target image. The network functions as a self-supervised regression network which uses final convolution layers in each path for generating deformation field components at different resolutions. There is no fully connected layer. The resulting deformations from each skip level are up-sampled to the same resolution as the input image, averaged, and used to warp desired regions of one of the input channels to corresponding regions of the other input. In the top, serial, path of the network, three convolution layers result in down sampled feature maps, followed by deconvolution layers which up-sample layer information back to the same resolution as the input image. The final convolution layer involves two channels which serve as row and column deformation components.

Optimization is performed using a cost function of the style in (1), using pixel-wise MAE as our similarity measure. We experimented with different forms of regularization. In essence, the network is self-supervised in that optimization is based solely on the similarity measure between the deformation-modified output image and the target image. As reported in [27], this type of network both learns and optimizes non-parametric deformation fields and can be applied to new data without retraining.

B. *Object-Level Registration*

Our approach for object level registration involves a custom architecture that involves fusing the results of two CNNs in order to find candidate matching pairs of lesions on the CC and MLO views. The architecture is shown in Figure 2. For each mammographic view, we use a CNN architecture known as Regions with CNN features (R-CNN) [11] which finds candidate lesions in the respective image. Whereas conventional CNNs process the entire image, R-CNNs first identify smaller candidate regions within the image and process only those regions. Hence, less memory is required which allows larger images to be processed—a characteristic that can be important for medical image processing where maintaining original pixel resolution can be important for preserving subtle differences in texture. For our R-CNN, we experimented with a custom region proposal function for finding candidate lesion areas. Further, we used a pre-trained network, which requires less training data, for the CNN component and fine-tuned it for the mammography problem.

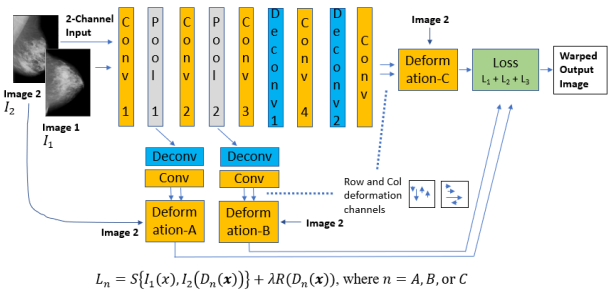

Figure 1: Pixel Level Registration CNN Architecture.

As shown in Figure 2, the output candidate regions from each R-CNN are paired and various combinations are formed. The score values provided by each R-CNN are used to filter out certain combinations of pairs. The remaining pairs are merged into a two-image-layer product. Yet we also create a third band, involving the absolute difference between relative distances-from-nipple for each detection as is shown in Figure 2. Hence, a three-image-layer input is provided to our final CNN, which classifies the pair candidates as matching or non-matching.

Somewhat related to this approach are multi-view CNN processing techniques, which have demonstrated that using a CNN to recover 3-D shapes from multiple 2-D images provided better performance than with 3-D representations [35]. Our approach was also motivated by an observation that lesions in the CC and MLO views showed a somewhat linear correlation in the relative lesion-to-nipple distance for the same lesions in both views as shown in Figure 3 for the lesions in CC and MLO images of the Curated Breast Imaging Subset of the Digital Database for Screening Mammography (CBIS-DDMS) [36] [37] [38].

IV. EXPERIMENTAL RESULTS

A. *Datasets*

We used two data sets for our experiments. The first is a synthetic, surrogate data set, involving 3-D versions of Mixed National Institute of Standards and Technology (MNIST) letters [39] which we generated using the package, iso2Mesh [40]. We use this set for validating the premise and

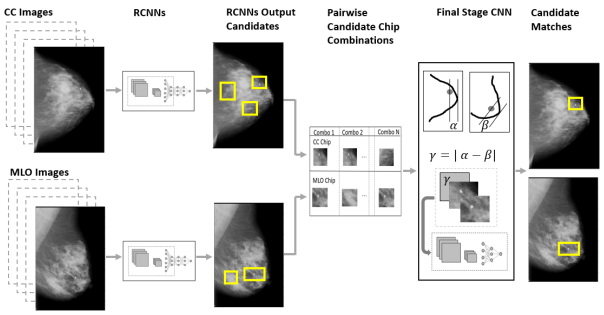

Figure 2: Object level Registration CNN Architecture.

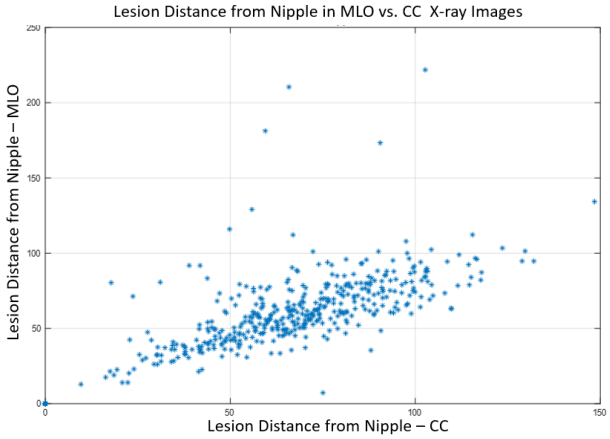

Figure 3: Plot of Lesion Distance to Nipple for MLO vs. CC images in CBIS-DDSM data set.

functionality of our algorithms. The second involves 2-D Scanned Film mammography images from the CBIS-DDSM data set. The CBIS-DDSM data set includes matched pairs of CC and MLO mammogram views, along with binary mask images that denote the location of lesions in the breast tissue. CBIS-DDSM is a publicly available mammography database [36]. Therefore, our experiments did not involve human subjects.

B. Pixel Level Registration

We first discuss experiments with our surrogate 3-D MNIST data set. These experiments involved using a simplified imaging model for simulating the projection from 3-D to 2-D views so as to evaluate the basic feasibility of the breast image registration problem.

The experiment involved the CNN illustrated in Figure 1, yet for the inputs we used two 2-D projections of 3-D MNIST letters taken from a 3-D object space in which we placed the 3-D letters. Figure 4 shows an example 3-D MNIST letter generated and the projection of three such letters into 2-D views. Each letter is randomly oriented, and also randomly

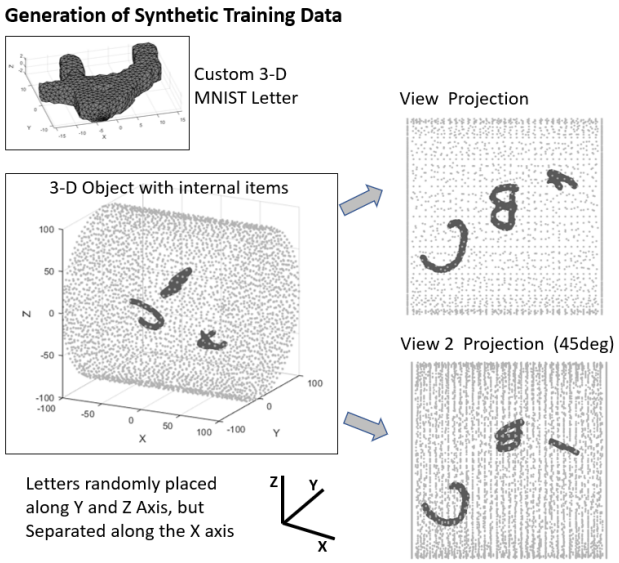

Figure 4: 3-D to 2-D Projection of 3-D MNIST letters.

positioned along the Y and Z axes of a cylindrical tube. The only non-random placement is along the X-axis, where we constrain a single letter in one of three bins along the X-axis. Some random overlap of letters among the bins is allowed. We generated the 2-D projections by saving a view of the cylinder at a given position, then rotated the entire cylinder space approximately 45 degrees, with some random variability and saving a second view. The resulting two projections serve as our two-channel input to the network shown in Figure 1. The images are 1024 x 1024 pixels, though we also tested lower resolutions such as 128 x 128 pixels. We generated approximately 9,000 instances of these random projection pairs (18,000 images total), and trained on 90% of them, using the remainder for testing. While this model is very simplistic in that it is rigid and does not simulate gray level texture of the breast nor the distortion effects of compression, it does provide a basis for testing the basic concept.

The results on an example test image in Figure 5 show the degree to which the network is able to warp the moving image such that it matches the target image. The network performed in this fashion in general on test data. The deformation field in Figure 5 shows the left-most side of the images (boxed in blue in the "Image to Register" panel), revealing pixel movement for the letters A and B.

Quantitative performance was assessed for several training and testing configurations. For all configurations, the mean intersection over union (IOU) [41] for binary versions of the registered vs. target images was consistently around 0.99 (on a scale of 0 to 1). For the gray scale versions of the images, normalized to intensity ranges between 0 and 1, the average Mean Squared Error (MSE) was approximately 0.0014 and average Peak Signal-to-Noise (PSNR) Ratio was 28. Hence, the warped images were nearly identical to the target images.

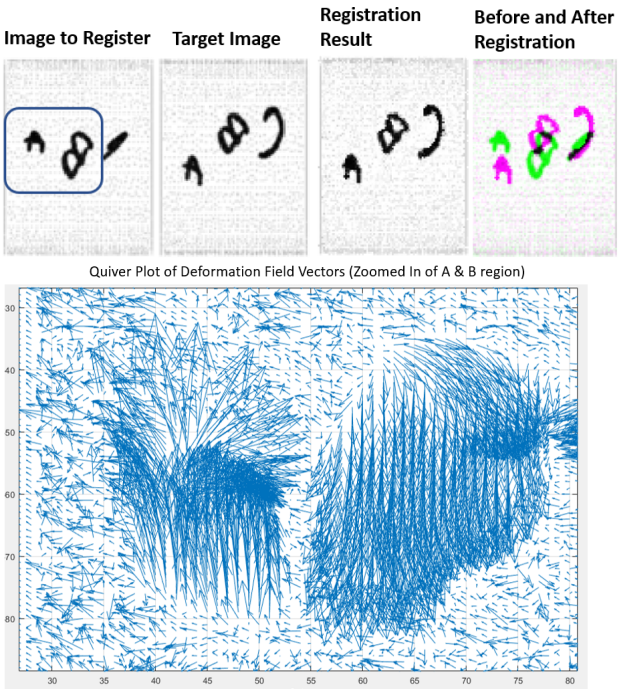

Figure 5: Pixel level Registration of two different MNIST letter projections and corresponding deformation field for a region.

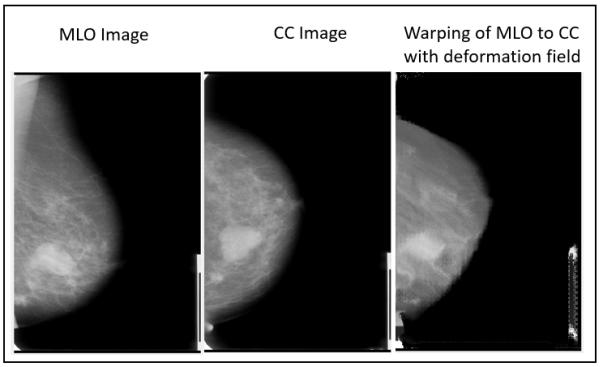

Figure 6: Pixel level registration results for CC and MLO X-ray images.

Our experiments with mammography images involved significantly less data (around 1000 image pairs with 90% used for training). Our network showed some ability to warp the images as is shown in Figure 6. However, it became evident that further modelling is needed in order to appropriately register the similar tissue areas. In essence, while the general form of our MLO image is warped to that of the CC target image, the texture and structure of the breast is not properly registered. We believe that there are several reasons for this (in addition to limited training data) such as the one-to-many pixel mappings and occlusions of tissue. Yet we believe that with sufficient training data and further refinements, such as an improved cost function, the network will be able to learn and match certain common breast tissue features, such as the brighter regions in the breasts in Figure 6. Further, it may be more feasible to test this type of network on 3-D DBT CC and MLO data, where there is a greater possibility of finding matching tissue and anatomical regions due to the full breast coverage of this data. Hence, given these observations, we hope to continue exploring the use of this type of network.

We considered ways of quantitatively assessing performance for the pixel based mammographic registration but it became evident that given the variability of breast tissue features, additional clinical insight is needed in order for such metrics to be meaningful. Hence, we plan to investigate this further from both a machine learning and clinical perspective.

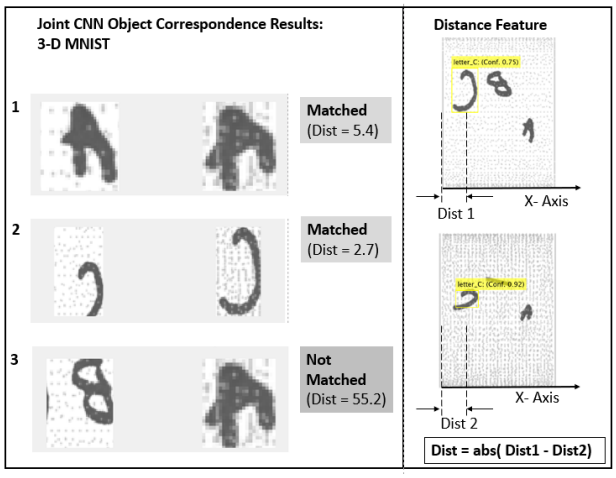

Figure 7: Object level registration (matching) of 3-D MNIST letters.

C. *Object Level Correspondence*

In this set of experiments, we sought to determine if a dual CNN and fusion based architecture, as shown in Figure 2, could achieve object level correspondence by correctly detecting and matching objects present in two image views. The same data sets from the pixel registration experiments were used here.

The R-CNN training for each of the dual paths involved fine-tuning an existing network that had already been trained on the CIFAR10 data set [42]. We used a minibatch size of 64 both for the MNIST data and X-ray data. The third CNN, in the final, fused, stage involved two convolutional layers (including ReLu and Max-Pooling) and a fully connected layer. A minibatch size of 64 was also used for training this CNN. For each data set, a third of the data was used for training the first-stage R-CNNs, another third for training the final-stage CNN, and the final third used to test the final-stage CNN. As noted in section III, combinations of image pairs are formed using the R-CNN output regions and this results in significantly more datapoints for the final CNN stage of the network.

The left panel in Figure 7 shows the results for three example MNIST pairs. As indicated, output pairs 1 and 2 were correctly predicted as matching. Output 3 (for letters B and A) were also correctly predicted as not matching. As noted earlier, an extra feature that is provided for the final stage CNN is a relative distance measure for each candidate object detected from the first stage R-CNNs. The right panel of Figure 7 illustrates the computation of this distance for the MNIST letters. In the left panel, this value is shown for each of the three example output pairs. The comparatively small values for outputs 1 and 2 reflect the fact that these two objects, in their respective original image projections were at approximately the same position along the X-axis, referring to our model in Figure 4. Recall that the 3-D cylinder space has been rotated about the X-axis. Therefore, while the objects from each projection will both appear different and while their positions along the plane of rotation may vary, their positions along the X-axis will remain the same. This roughly simulates the notion of the relative distance of a breast lesion from the nipple as it appears between the CC and MLO mammographic views.

Quantitatively, the final stage of the network was tested on 33,909 image pairs resulting from pair-wise combinations of regions detected by each R-CNN from the first stage. Of these, 10,683 were actual matching pairs (e.g., A and A) and the network accurately predicted 98.56% of these as matching (10,529). For the remaining 23,226 non-matching candidates, the network accurately predicted 96.77% as non-matching.

For our experiments with this network using mammographic data, the results showed significant potential visually, but varied. Figure 8 shows one of the best performing output cases in which a single lesion (based on the true data, indicated by the blue boxes) was both detected and matched between the CC and MLO views. First stage R-CNN scores for the lesions were 0.85 (CC) and 0.74 (MLO). However, Figure 9 shows a more commonly observed output case in which the correct lesion was matched, but with significant false alarms. Yet, it is noted that for this particular case, the R-CNN scores for the correctly matched lesions were 0.97 (CC)

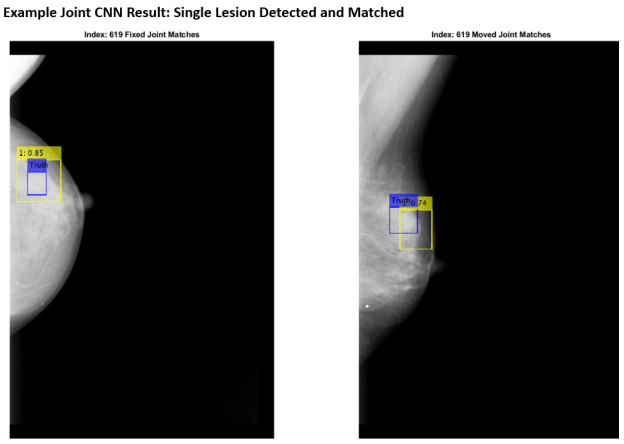

Figure 8: Results of object level CNN registration of lesions in CC and MLO X-ray images.

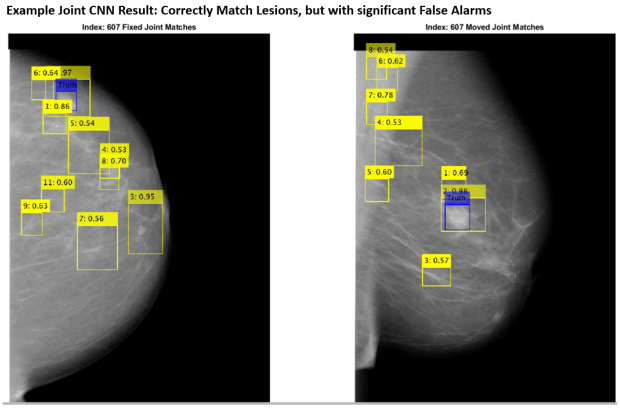

Figure 9: Results of object level CNN registration of lesions in CC and MLO X-ray images (second example).

and .98 (MLO), whereas the scores for most of the false alarms were less than 0.7. Hence, false alarms could be filtered out using score thresholds or other information.

As similarly discussed for our pixel-based registration technique, we plan to further investigate, with additional clinical insight, appropriate quantitative metrics for the object level correspondence of the mammography data.

While the object level correspondence results are preliminary, we believe that with additional training and refinement, this network architecture can be improved to provide reasonable performance. Here again, potential improvements may also be realized by using 3-D DBT data, which would allow for more localized tissue information to be provided to the network.

V. Conclusion

We have explored techniques for performing automated registration for mammographic images taken from multiple views. We performed preliminary experiments with two CNN architectures, involving custom modifications, for this problem. One architecture involved a non-rigid deformation field-based pixel level registration technique. The other involved a dual-path, fusion-based architecture for performing object level feature correspondence. We use a surrogate, rigidly-transformed data set involving 3-D MNIST letters for simulating certain basic geometric modelling facets of the breast imaging process. Moreover, we also utilized real mammographic image data from the CBIS-DDSM data set.

Our results showed that both networks could successfully register the MNIST-based data from both a visual and quantitative perspective. However, challenges were revealed for registering or performing objective correspondence with the X-ray mammography data. Still, both networks showed areas of potential which suggest that, with modifications, improvements could be made. Observed areas to address for the deformation-field based CNN network included utilizing a more appropriate similarity measure and applying an appropriate regularization technique. However, it is also assessed that this type of network may perform better using more 3-D-like DBT mammographic images, where it is possible that more one-to-one pixel (or local region) type mappings may be possible. For the joint-fusion based network, additional training, additional data, experimentation with different pre-trained CNNs (or custom trained CNNs), and the use of additional image statistical features are areas that should yield improvements. For both techniques we plan to further investigate, with additional clinical insight, ways of quantitatively characterizing registration performance for the mammography data and means of improving the networks in general, along with identifying additional potential use cases.

References

[1] American Cancer Society, "How common is breast cancer?," 2019. [Online]. Available: https://www.cancer.org. [Accessed 19 1 2019].

[2] R. L. Siegel, K. D. Miller and A. Jemal, "Cancer statistics, 2015," *CA: A Cancer Journal for Clinicians,* vol. 65, no. 1, pp. 5-29, 2015.

[3] National Breast Cancer Foundation, Inc., "About breast cancer," 2016. [Online]. Available: https://www.nationalbreastcancer.org/diagnostic-mammogram. [Accessed 20 1 2016].

[4] Radiological Society of North America, "RadiologyInfo.org," 2019. [Online]. Available: https://www.radiologyinfo.org/en/info.cfm?pg=mammo#overview. [Accessed 16 1 2019].

[5] L. C. Miller, "Mammography positioning: basic and advanced," 2016. [Online]. Available: https://www.sbi-online.org. [Accessed 20 1 2019].

[6] D. Porter, W. Walton, S. Harvey, B. Tsui, K. Peyton, M. Jones, T. Lee, P. Yi, F. Hui, H. Sair, E. Ambinder, L. Mullen and S.-J. Kim, "Multimodality machine learning for breast cancer detection: Synergistic performance with upstream Data Fusion of digital breast tomosynthesis and ultrasound," in *Machine Learning for Health Care*, Stanford, CA, August, 2018.

[7] J. Guo, R. Sivaramakrishna, C.-C. Lu, J. S. Suri and S. Laxminarayan, "Breast image registration techniques: a survey," *Medical and Biological Engineering and Computing,* vol. 44, no. 1-2, pp. 15-26, 2006.

[8] J. H. Hipwell, C. Tanner, W. R. Crum, J. A. Schnabel and D. J. Hawkes, "A new validation method for X-ray mammogram registration algorithms using a projection model of breast X-ray compression," *IEEE Transactions on Medical Imaging,* vol. 26, no. 9, 2007.

[9] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. v. d. Laak, B. v. Ginneken and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Medical Image Analysis,* vol. 42, pp. 60-88, 2017.

[10] J.-G. Lee, S. Jun, Y.-W. Cho, H. Lee, G. B. Kim, J. B. Seo and N. Kim, "Deep learning in medical imaging: General overview," *Korean Journal of Radiology,* vol. 18, pp. 570-584, 2017.

[11] R. Girshick, J. Donahue, T. Darrell and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. CVPR*, 2014.

[12] F. Alam and S. U. Rahman, "Intrinsic registration techniques for medical images: A state-of-the-art review," *Journal of Postgraduate Medical Institute,* vol. 30, no. 2, 2016.

[13] F. P. M. Oliveira and J. M. R. S. Tavares, "Medical image registration: a review," *Computer Methods in Biomechanics and Biomedical Engineering,* vol. 17, no. 2, pp. 73-93, 2014.

[14] R. Shams, P. Sadeghi, R. A. Kennedy and R. I. Hartley, "A survey of medical image registration on multicore and the GPU," *IEEE Signal Processing Magazine,* vol. 27, no. 2, 2010.

[15] L. G. Brown, "A survey of image registration techniques," *ACM Computing Surveys,* vol. 24, no. 4, 1992.

[16] C. Wachinger, "MICCAI 2010 tutorial: Intensity-based deformable registration - similarity measures," 2010. [Online]. Available: http://campar.in.tum.de/DefRegTutorial/WebHome. [Accessed 1 1 2019].

[17] N. Ruiter, T. Muller, R. Stotzka and W. Kaiser, "Elastic registration of x-ray mammograms and three-dimensional magnetic resonance imaging data," *Journal of Digital Imaging,* vol. 14, no. 2, pp. 52-55, 2001.

[18] N. Ruiter, T. Muller, R. Stotzka, H. Gemmeke, J. Reichenbach and W. Kaiser, "Automatic image matching for breast cancer diagnostics by a 3D deformation model of the mamma," *Biomedical Engineering,* vol. 47, no. s1b, pp. 644-647, 2009.

[19] T. Hopp, M. Dietzel, P. Baltzer, P. Kreisel, W. Kaiser, H. Gemmeke and N. Ruiter, "Automatic multimodal 2D/3D breast image registration using biomechanical FEM models and intensity-based optimization," *Medical Image Analysis,* vol. 17, pp. 209-218, 2013.

[20] M. Samulski and N. Karssemeijer, "Matching mammographic regions in mediolateral oblique and cranio caudal views: a probabilistic approach," in *Proc. SPIE Medical Imaging 2008*, 2008.

[21] Y. Kita, R. Highnam and M. Brady, "Correspondence between different view breast X rays using curved epipolar lines," *Computer Vision and Image Understanding,* vol. 83, no. 1, pp. 38-56, 2001.

[22] Y. Kita, R. Highnam and M. Brady, "Correspondence between different view breast X-rays using a simulation of breast deformation," in *Proc. of the 1998 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Santa Barbara, CA, USA, 1998.

[23] J. Krebs, T. Mansi, H. Delingette, L. Zhang, F. C. Ghesu, S. Miao and A. K. Maier, "Robust non-rigid registration through agent-based action learning," in *Proc. International Conference on Medical Image Computing and Computer-Assisted Intervention*, Quebec City, Canada, 2017.

[24] B. D. d. Vos, F. F. Berendsen, M. A. Viergever, M. Staring and I. Išgum, "End-to-end unsupervised deformable image registration with a convolutional neural network," in *Proc. International Workshop on Deep Learning in Medical Image Analysis*, Quebec City, Canada, 2017.

[25] H. Sokooti, B. d. Vos, F. Berendsen, B. P. F. Lelieveldt, I. Išgum and M. Staring, "Nonrigid image registration using multi-scale 3D convolutional neural networks," in *Proc. International Conference on Medical Image Computing and Computer-Assisted Intervention*, Quebec City, Canada, 2017.

[26] I. Yoo, D. G. C. Hildebrand, W. F. Tobin, W.-C. A. Lee and W.-K. Jeong, "ssEMnet: Serial-section electron microscopy image registration using a spatial transformer network with learned features," in *Proc. International Workshop on Deep Learning in Medical Image Analysis*, Quebec City, Canada, 2017.

[27] H. Li and Y. Fan, "Non-rigid image registration using self-supervised fully convolutional networks without training data," in *Proc. of the IEEE International Symposium on Biomedical Imaging 2018 (ISBI 18)*, Washington, D.C., 2018.

[28] K. A. J. Eppenhof, M. W. Lafarge, P. Moeskops, M. Veta and J. P. W. Pluim, "Deformable image registration using convolutional neural networks," in *Proc. SPIE 10574, Medical Imaging 2018: Image Processing*, Houston, Texas, 2018.

[29] X. Yang, R. Kwitt, M. Styner and M. Niethammer, "Quicksilver: Fast predictive image registration – A deep learning approach," *NeuroImage,* vol. 158, pp. 378-396, 2017.

[30] S. Perek, A. Hazan, E. Barkan and A. Akselrod-Ballin, "Mammography dual view mass correspondence," in *Proc. Computer Vision and Pattern Recognition*, 2018.

[31] D. Zikic, "MICCAI 2010 tutorial: Intensity-based deformable registration - Overview of intensity-based deformable registration," 24 9 2010. [Online]. Available: http://campar.in.tum.de/DefRegTutorial/WebHome. [Accessed 23 1 2019].

[32] M. Staring, "MICCAI 2010 tutorial: Intensity-based deformable registration - regularization in deformable registration," 24 9 2010. [Online]. Available: http://campar.in.tum.de/DefRegTutorial/WebHome. [Accessed 23 1 2019].

[33] V. Vishnevskiy, T. Gass, G. Szekely, C. Tanner and O. Goksel, "Total variation regularization of displacements in parametric image registration," in *Proc. International MICCAI Workshop on Computational and Clinical Challenges in Abdominal Imaging*, Cambridge, MA, 2014.

[34] J. Long, E. Shelhamer and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, Massachusetts, 2015.

[35] H. Su, S. Maji, E. Kalogerakis and E. Learned-Miller, "Multi-view convolutional neural networks for 3D shape recognition," in *Proc. IEEE International Conference on Computer Vision*, Santiago, Chile, 2015.

[36] R. S. Lee, F. Gimenez, A. Hoogi and D. Rubin, "Curated breast imaging subset of DDSM," The Cancer Imaging Archive, 2016.

[37] R. S. Lee, F. Gimenez, A. Hoogi, K. K. Miyake, M. Gorovoy and D. L. Rubin, "A curated mammography data set for use in computer-aided detection and diagnosis research," *Scientific Data,* vol. 4, 2017.

[38] K. Clark, B. Vendt, K. Smith, J. Freymann, J. Kirby, P. Koppel, S. Moore, S. Phillips, D. Maffitt, M. Pringle, L. Tarbox and F. Prior, "The cancer imaging archive (TCIA): Maintaining and operating a public information repository," *Journal of Digital Imaging,* vol. 26, no. 6, pp. 1045-1057, 2013.

[39] National Institute of Standards & Technology, "NIST Special Database 19," National Institute of Standards & Technology, 27 8 2010. [Online]. Available: https://www.nist.gov/srd/nist-special-database-19. [Accessed 20 1 2019].

[40] Q. Fang and D. Boas, "Tetrahedral mesh generation from volumetric binary and gray-scale images," in *Proc. of the IEEE International Symposium on Biomedical Imaging*, 2009.

[41] A. A. Taha and A. Hanbury, "Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool," *BMC Medical Imaging,* vol. 19, 2015.

[42] A. Krizhevsky, "Learning multiple layers of features from tiny images," 2009.