Motion Corruption Detection in Breast DCE-MRI

Sylvester Chiang[1(✉)], Sharmila Balasingham[2], Lara Richmond[2], Belinda Curpen[2], Mia Skarpathiotakis[2], and Anne Martel[1]

[1] Department of Medical Biophysics, University of Toronto, Toronto, ON, Canada
sylvester.chiang@mail.utoronto.ca
[2] Sunnybrook Health Sciences Centre, Imaging Research, Toronto, ON, Canada

Abstract. Motion corruption can result in difficulty identifying lesions, and incorrect diagnoses by radiologists in cases of breast cancer screening using DCE-MRI. Although registration techniques can be used to correct for motion artifacts, their use has a computational cost and, in some cases can lead to a reduction in diagnostic quality rather than the desired improvement. In a clinical system it would be beneficial to identify automatically which studies have severe motion corruption and poor diagnostic quality and which studies have acceptable diagnostic quality. This information could then be used to restrict registration to only those cases where motion correction is needed, or it could be used to identify cases where motion correction fails. We have developed an automated method of estimating the degree of mis-registration present in a DCE-MRI study. We experiment using two predictive models; one based on a feature extraction method and a second one using a deep learning approach. These models are trained using estimates of deformation generated from unlabeled clinical data. We validate the predictions on a labeled dataset from radiologists denoting cases suffering from motion artifacts that affected their ability to interpret the image. By calculating a binary threshold on our predictions, we have managed to identify motion corrupted cases on our clinical dataset with an accuracy of 86% based on the area under the ROC curve. This approach is a novel attempt at defining a clinically relevant level of motion corruption.

Keywords: DCE-MRI · Motion artifacts · Registration · Feature extraction · Convolutional neural networks

1 Introduction

Dynamic contrast-enhanced MRI (DCE-MRI) has been shown to increase the sensitivity of cancer detection in breast screening programs for women at higher risk of breast cancer [3]. In these exams, a single pre-contrast image and a series of post-contrast images are acquired and the degree and rate of contrast-enhancement is used to identify cancer. Motion corruption can create artifacts in subtraction images which may obscure lesions and decrease the diagnostic utility of the images. Automated non-rigid registration [6] can be used to align temporally sequential sequences of DCE-MRI and are commonly used as a pre-processing step in computer aided

Q. Wang et al. (Eds.): MLMI 2017, LNCS 10541, pp. 10–18, 2017.
DOI: 10.1007/978-3-319-67389-9_2

detection algorithms [9]. However there are unique difficulties in registering DCE-MRI images, including the non-rigid deformability of breast tissue, lack of clear landmarks, and the presence of contrast-enhancement. These characteristics can result in the failure of registration resulting in a reduction in image quality [2]. Furthermore, the use of image registration inevitably results in some blurring due to interpolation [2] and this is particularly a problem in clinical breast imaging where the pixels are very anisotropic. Registration also has an associated time penalty with non-rigid registration algorithms running on the order of a single minute to upwards of 15 min [2].

This paper proposes an automated method for motion-corruption detection in DCE-MRI scans. This would allow for an independent assessment of registration quality that is completely separate from any optimization metric in the registration process. This predictive model could also serve in a CAD pipeline to determine the necessity of registration, thereby reducing the associated time penalty and removing the possibility of mis-registration. A unique challenge associated with this task is to identify motion artifacts that are clinically relevant and need to be corrected to allow for proper interpretation. To build this binary model, we first attempted to use a feature extraction approach that has been used for registration quality assessment in other imaging modalities and biological models [1, 5]. We also utilized a deep learning approach to generate binary labels using convolutional neural networks to draw a comparison between the different learning paradigms.

2 Methodology

We propose a two-stage process to detect motion corruption. First, we estimated the average deformation using a regression model trained on unlabeled data. Then we used the output from the regression models to predict a binary categorical label that corresponded to the motion-corruption labels provided by radiologists. We explored supervised learning methods using a traditional feature extraction approach and drew comparisons with a neural network model. Receiver operating characteristic (ROC) curves allowed us to analyze the trade-off between the true positive rate (TPR) and false positive rate (FPR) at different binary thresholds making it more clinically adaptable (see Fig. 1).

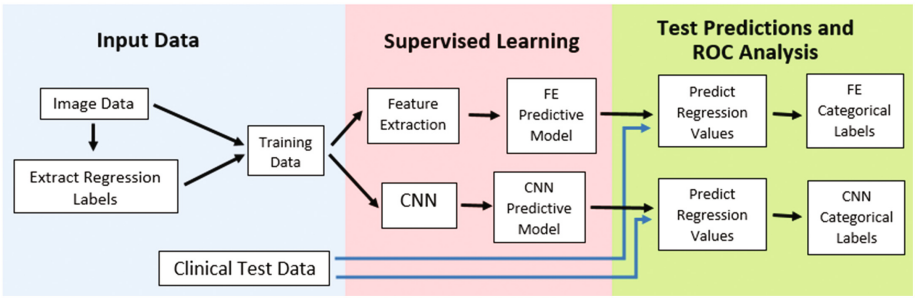

Fig. 1. This is the learning model for our proposed experiment. Regression labels were extracted from the DCE-MRI data and used to train two separate supervised learning models. The test set was then fed to both models to predict regression values which were used to infer a binary label.

2.1 Data Acquisition

In our experiments we work with DCE-MRI data which is acquired at our institution as a set of volumes consisting of a pre-contrast volume and four post-contrast volumes. The post-contrast volumes are acquired approximately 90 s apart to allow for circulation of the contrast agent. The dimension of each volume was $512 \times 512 \times 80$ with an in-plane resolution of approximately 0.35 mm and a slice thickness of 3 mm. Subtraction images were generated by subtracting the pre-contrast image from each of the post-contrast images with each subtraction treated as a separate instance.

To obtain a training dataset we pulled unlabeled volumes from our database of breast MR images and associated radiology reports. Reports with motion-identifying phrases such as "motion corrupted" and "motion artifacts" were queried to collect a sample of motion-corrupted volumes. We then conducted another search filtering out these phrases and selected an equivalent number of volumes to acquire a set of data free of severe motion-corruption. In total we obtained a balanced dataset of 88 DCE-MRI volumes. To validate our results, we acquired a completely separate set of 40 patient MRIs to use as our final test set. These volumes were provided by radiologists asked to submit motion-corrupted cases they recognized in their daily work. To keep a balanced dataset they were requested to submit a case free of motion-related artifacts in conjunction with the motion-corrupted images. There was a high degree of variation within each class, as individual radiologists tolerated different amounts of motion.

2.2 Generating Deformation Estimates for Unlabeled Images

We could not assume that the absence of a comment about motion corruption in the radiology report was sufficient to label an image as being free of motion-corruption. This meant that it was necessary to generate surrogate labels to quantify the deformation present in these images for the purpose of supervised learning. We used the Elastix library to perform 3D b-spline registration on these images and then used Transformix to output the resulting deformation fields. The field contained vectors dictating the pixel shift for the registration, and the Euclidean distance for each vector was calculated and averaged over the entire volume to calculate a single metric. This process allowed us to capture the quantity of deformation in a volume as a continuous variable which we could then use to train a regression model via supervised learning.

2.3 Preprocessing Data

The feature extraction method was performed on a set of four inputs; the pre-contrast image, post-contrast image, subtraction image, and an associated phase correlation map. Phase correlation uses a frequency domain approach to estimate the translational offset between two images [7] and is robust in quantifying translation even in the presence of contrast agents. In accordance with the Fourier shift property the spatial translation is represented by a phase change in the spectral domain; this phase property is then calculated by identifying peaks in the cross-correlation between sub-regions. The map was generated by iterating through patches of the pre-contrast and post-contrast image to calculate a dense phase correlation map.

In our experiments, only one set of pre-contrast and post-contrast images were considered at a time. Different institutions utilize different imaging protocols and may produce a different number of post-contrast images. Creating a model to only use one pre-contrast and one post-contrast image would be more generalizable across institutions and could be more easily adopted. Segmentations were applied to the breast volumes to remove the surrounding air and chest wall which were not significant in the analysis of breast deformation and could have in fact added more irrelevant noise. The DCE-MRI images were acquired as 3D volumes but slices were extracted along the center of each breast to augment the number of instances in our training model.

2.4 Training a Supervised Learning Model

The regression models discussed were trained on the dataset of 88 patients. We augmented our dataset by taking slices from each 3-dimensional volume, in total producing 3180 training instances. This image set was split into an 80% training set and 20% tuning set, with data stratification on a volume level. We compared two different feature extraction approaches in our experiments. The first was based on hand-crafted features that were then used as an input vector for a variety of classical machine learning algorithms. A second approach consisted of utilizing a Convolutional Neural Network (CNN) based on deep learning theory. This type of network contains convolutional layers which act as high-level feature extractors that operate on raw images as input. These extracted features are then fed into a series of dense layers for classification.

Feature Extraction. Features were calculated globally, within the inner volume of the breast, and the edges of the breast near the skin because motion artifacts that occur in particular regions may be more significant. These region masks were generated by eroding segmentations. 15 intensity-based features were calculated for each region, including mean, mode, skew, and negative sum of squares (NSSQ) which was the sum of squares exclusively calculated over negative regions. These features were computed over the subtraction images while mutual information features were calculated using the pre-contrast and post-contrast images.

Texture analysis using gray-level co-occurrence matrices provided a second source of features. Texture measures are highly utilized in computer vision problems within the medical field. This extraction method examined the contrast, correlation and homogeneity of images. These texture features were also applied to the phase correlation maps. In total we generated a set of 52 features from each 2D sample. These features were used as input to machine learning algorithms available through sklearn, specifically Support Vector Machines (SVMs) and tree-based ensemble learning methods.

Convolutional Neural Networks. Deep learning networks are computational models built to learn hierarchal representations of data at different levels of abstraction CNNs have been applied to image classification tasks with a high amount of success, achieving record accuracies [4]. As opposed to dense mappings between individual perceptrons within the hidden layers, CNNs learn a large number of filters that are convolved with a 2D input data. These networks perform well on computer vision

problems due to a reduced parameter space and their usage of translationally invariant filters. Another advantage of utilizing CNNs is that image data can generally be directly fed into the network with minimal pre-processing steps, unlike feature extraction which requires a manual exploration of discriminative metrics.

Our architecture was inspired by a simple CNN, AlexNet [4] with two reduced layers to make the network easier to train due to fewer parameters. It consisted of four convolutional layers followed by a dense mapping to two fully-connected layers (Fig. 2). We chose rectified linear units (ReLU) as our activation function, which are lightweight and helped to address the vanishing gradient problem [4]. Dropout was also added as a regularization tool and operates by masking random activations therefore preventing neighboring nodes from clustering together to learn specific representations of image features [8]. The input of the CNN was comprised of 3 channels; the pre-contrast, post-contrast, and phase correlation map which were all downsampled to 256 × 256 pixels. The neural networks were built using the Theano and Lasagne libraries in Python which allowed for optimized calculations to power the learning process (see Fig. 3).

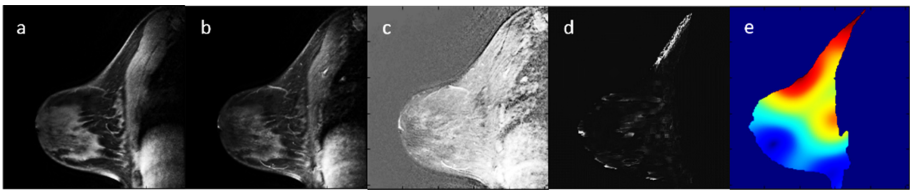

Fig. 2. Visualized input data, pre-contrast image (a), post-contrast image (b), subtraction image with a noticeable motion artifact along the top edge of the breast (c), segmented phase correlation map (d), segmented deformation field (e)

Inputs (3@256x256)
Convolutional Layer (32@128x128)
Convolutional Layer (48@64x64)
Convolutional Layer (96@32x32)
Convolutional Layer (96@16x16)
Fully-Connected Layer (2048 units)
Fully-Connected Layer (1024 units)

Fig. 3. The architecture of the CNN. The first convolutional layer is the product of a (5 × 5) kernel followed a (2 × 2) max-pool. The following three layers were convolved by a (3 × 3) filter again followed a (2 × 2) max-pool. The activations after the fourth convolutional layer were flattened and mapped to a dense fully-connected layer.

3 Results

3.1 Comparing Learning Models

Both supervised learning methods were trained on the set of 88 patients and then tested on the radiologist-labelled set of 40 DCE-MRI images. The regression values were used to generate thresholds and subsequently infer categorical labels. The sensitivity (TPR) and specificity (1-FPR) at each threshold value were plotted to form an ROC curve representing the accuracy of the predictions compared to our ground truth labels. Averaging the deformation predictions of individual slices to generate a volume-level prediction before constructing the ROC increased the AUC value. The increase in aggregate mean accuracy was likely due to the labelling being performed on a volume basis. Taking the mean value offered a better representation of the 3D volume and as a result predicted the ground truth more accurately. However the results from the hand-crafted features still underperformed compared to the CNN (see Fig. 4).

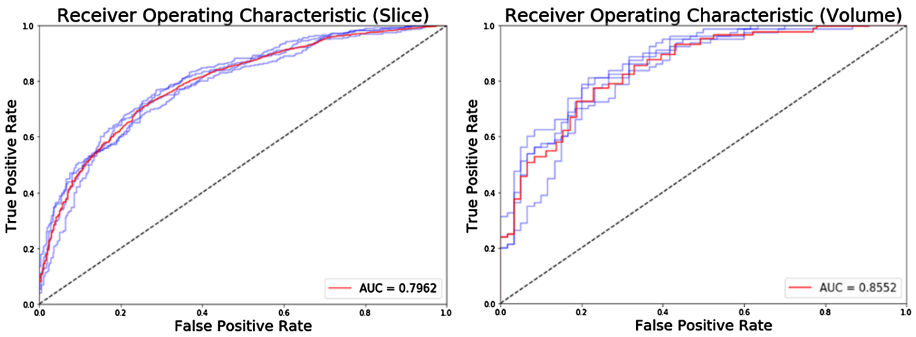

Fig. 4. The ROC curve calculated from the regression values predicted by the CNN. Blue lines represent different randomly initialized neural networks and the red represents the average ROC. The left image shows the curve constructed for every slice and the right curve plots the deformation metric aggregated over volumes (6 slices). (Color figure online)

The same training and testing paradigm was used on the CNN model. The AUC of ROC value was 0.80 ± 0.01 and when averaged over volumes demonstrated an AUC value of 0.86 ± 0.01 (Fig. 5). CNNs achieved a predictive power higher than previous supervised learning methods based on hand-crafted feature extraction. On a 3.6 GHz CPU, phase map preprocessing required on average 5.4 s followed by a forward pass of this CNN which took approximately 0.5 s. These values compared favorably to the Elastix registration time which ran for 140 s per volume on average. The advantage of running this classification model could yield a large potential temporal benefit as this time penalty is incurred for every post-contrast image leading to large processing times (see Table. 1).

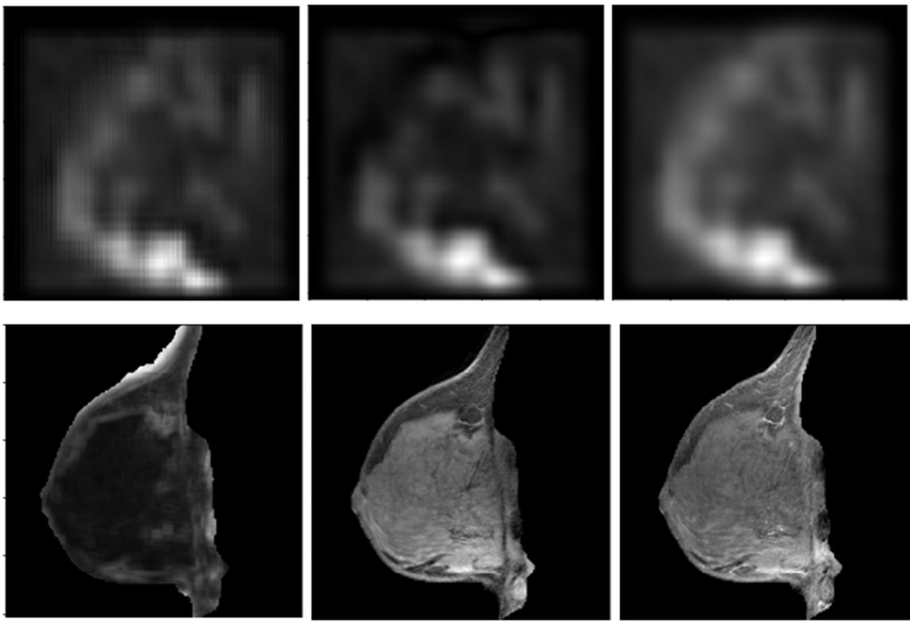

Fig. 5. The top row shows the maximum activations in the first dense layer of the network deconvolved, and the corresponding input images shown on the bottom row. From left to right columns, the images show the phase map, the pre-contrast image, and the post-contrast image.

Table 1. A table of the AUC metric calculated on ROC curves. The left column is calculated for each individual slice and the right column shows the metric per volume.

AUROC (area under the receiver operator characteristic curve)		
Learning method	Per slice	Aggregate mean
Random forest (n = 100)	0.65 ± 0.02	**0.71** ± 0.02
Random forest (n = 500)	0.65 ± 0.02	**0.71** ± 0.02
AdaBoost (n = 100)	0.69 ± 0.03	**0.72** ± 0.03
AdaBoost (n = 500)	0.65 ± 0.03	**0.70** ± 0.05
SVM	0.72 ± 0.01	**0.77** ± 0.01
CNN	0.80 ± 0.01	**0.86** ± 0.01

3.2 Visualizing Neural Networks

While neural network features can be difficult to understand, new visualization methods can help provide visual cues of what the learned features represent. Deconvolution nets project large activations back to the input space to better understand visual features that strongly impact the final classification score. We extracted the maximum activations in various layers and applied a deconvolution using the set of transposed weights learned from the forward pass. This was followed by an unpooling operation [10] to regain an

output of the same dimension as the input. The deconvolved images showed that large activations have correspondence with features at the edge of the breast. This provided evidence that even though features internal to the breast had a role in determining the amount of motion, the significant motion artifacts that are noted by clinicians tended to occur at the surface of the breast.

4 Conclusion

In this paper we compared the classical method of utilizing feature extraction to solve classification problems, to more recent deep learning approaches in the domain of registration accuracy. We achieved an accuracy of 0.86 based on the AUC of the ROC curve. We theorize that hand-crafted features rely on intuition and are often influenced by features common throughout the field of medical imaging. CNNs differ by extracting extremely domain-specific features at deeper convolution levels that are represented as a pattern of weights resulting in higher classification performance.

Motion corruption detection in DCE-MRI volumes remains a challenging task but automated assessment of these cases will become more impactful as larger amounts of medical information are processed. In this experiment we explored a preliminary method to predict motion corruption as it pertains to clinical interpretation. The model that we proposed is robust to contrast-enhancement due to the inclusion of phase correlation maps which register translation on sub-regions. We also introduced a method of estimating deformation in an image that is independent of physician subjectivity, and did not require any manual creation of landmarks. Finally we have developed a model to define a level of motion corruption that is relevant to a clinician's ability to diagnose an image, allowing for a more intelligent approach to applying registrations.

References

1. Armato, S.G., et al.: Temporal subtraction in chest radiography: automated assessment of registration accuracy. Med. Phys. **33**(5), 1239 (2006)
2. Klein, A., et al.: Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. Neuroimage **46**(3), 786–802 (2009)
3. Kok, T., et al.: Efficacy of MRI and Mammography for breast-cancer screening in women with a familial or genetic predisposition. Engl. J. Med. **351**(5), 427–437 (2004)
4. Krizhevsky, A., et al.: ImageNet classification with deep convolutional neural networks. Adv. Neural Inf. Process. Syst. 1–9 (2012)
5. Muenzing, S.E.A., et al.: Supervised quality assessment of medical image registration: application to CT lung registration. Med. Image Anal. **16**(8), 1521–1531 (2012)
6. Rueckert, D., et al.: Nonrigid registration using free-form deformations: application to breast MR images. IEEE Trans. Med. Imaging **18**(8), 712–721 (1999)
7. Srinivasa Reddy, B., Chatterji, B.N.: An FFT-based technique for translation, rotation, and scale-invariant registration. IEEE Trans. Image Process. **5**(8), 1266–1271 (1996)

8. Srivastava, N., et al.: Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. **15**, 1929–1958 (2014)
9. Vignati, A., et al.: A fully automatic lesion detection method for DCE-MRI fat-suppressed breast images. SPIE Med. Imaging **7260**, 726026 (2009)
10. Zeiler, M.D., Fergus, R.: Visualizing and understanding convolutional networks. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014. LNCS, vol. 8689, pp. 818–833. Springer, Cham (2014). doi:10.1007/978-3-319-10590-1_53

http://www.springer.com/978-3-319-67388-2

Machine Learning in Medical Imaging
8th International Workshop, MLMI 2017, Held in Conjunction with MICCAI 2017, Quebec City, QC, Canada, September 10, 2017, Proceedings

Wang, Q.; Shi, Y.; Suk, H.-I.; Suzuki, K. (Eds.)

2017, XV, 391 p. 134 illus., Softcover
ISBN: 978-3-319-67388-2