Automated Deep Learning Method for Whole-Breast Segmentation in Diffusion-Weighted Breast MRI

Lei Zhang, PhD[1], **Aly A. Mohamed, PhD**[1], **Ruimei Chai, MD**[1,2], **Yuan Guo, MD**[1,3], **Bingjie Zheng, MD**[1,4], **Shandong Wu, PhD**[5,*]

[1]Department of Radiology, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania, USA

[2]Department of Radiology, First Hospital of China Medical University, Heping District, Shenyang, Liaoning, China

[3]Department of Radiology, Second Affiliated Hospital of South China University of Technology, Guangzhou First People's Hospital, Guangzhou, China

[4]Department of Radiology, Henan Cancer Hospital, Affiliated Cancer Hospital of Zhengzhou University, Zhengzhou, Henan, China

[5]Departments of Radiology, Biomedical Informatics, Bioengineering, Intelligent Systems, and Clinical and Translational Science, University of Pittsburgh, Pittsburgh, Pennsylvania, USA

Abstract

Background: Diffusion-weighted imaging (DWI) in MRI plays an increasingly important role in diagnostic applications and developing imaging biomarkers. Automated whole-breast segmentation is an important yet challenging step for quantitative breast imaging analysis. While methods have been developed on dynamic contrast-enhanced (DCE) MRI, automatic whole-breast segmentation in breast DWI MRI is still underdeveloped.

Purpose: To develop a deep/transfer learning-based segmentation approach for DWI MRI scans and conduct an extensive study assessment on four imaging datasets from both internal and external sources.

Study Type: Retrospective.

Subjects: In all, 98 patients (144 MRI scans; 11,035 slices) of four different breast MRI datasets from two different institutions.

Field Strength/Sequences: 1.5T scanners with DCE sequence (Dataset 1 and Dataset 2) and DWI sequence. A 3.0T scanner with one external DWI sequence.

Assessment: Deep learning models (UNet and SegNet) and transfer learning were used as segmentation approaches. The main DCE Dataset (4,251 2D slices from 39 patients) was used for pre-training and internal validation, and an unseen DCE Dataset (431 2D slices from 20 patients) was used as an independent test dataset for evaluating the pre-trained DCE models. The main DWI Dataset (6,343 2D slices from 75 MRI scans of 29 patients) was used for transfer learning

*Address reprint requests to: S.W., 3362 Fifth Ave., Pittsburgh, PA, 15213. wus3@upmc.edu.

and internal validation, and an unseen DWI Dataset (10 2D slices from 10 patients) was used for independent evaluation to the fine-tuned models for DWI segmentation. Manual segmentations by three radiologists (>10-year experience) were used to establish the ground truth for assessment. The segmentation performance was measured using the Dice Coefficient (DC) for the agreement between manual expert radiologist's segmentation and algorithm-generated segmentation.

Statistical Tests: The mean value and standard deviation of the DCs were calculated to compare segmentation results from different deep learning models.

Results: For the segmentation on the DCE MRI, the average DC of the UNet was 0.92 (cross-validation on the main DCE dataset) and 0.87 (external evaluation on the unseen DCE dataset), both higher than the performance of the SegNet. When segmenting the DWI images by the fine-tuned models, the average DC of the UNet was 0.85 (cross-validation on the main DWI dataset) and 0.72 (external evaluation on the unseen DWI dataset), both outperforming the SegNet on the same datasets.

Data Conclusion: The internal and independent tests show that the deep/transfer learning models can achieve promising segmentation effects validated on DWI data from different institutions and scanner types. Our proposed approach may provide an automated toolkit to help computer-aided quantitative analyses of breast DWI images.

Level of Evidence: 3

Technical Efficacy: Stage 2

BREAST CANCER is one of the most common cancers and the second leading cause of death among women worldwide. The incidence of breast cancer in the world, especially in developing countries, is growing.[1] Breast magnetic resonance imaging (MRI) plays an important role in high-risk women breast cancer screening and for clinical problem-solving purposes. While breast dynamic-contrast enhanced (DCE) MRI is being widely used in clinical settings, diffusion-weighted imaging (DWI) in breast MRI is also a useful sequence for many diagnostic applications and in developing novel imaging biomarkers. DCE MRI is considered "invasive" due to the administration of the contrast agent, and gadolinium deposits in brains have raised safety concerns.[2] Different from DCE, DWI assesses how freely water molecules can diffuse within tissue, and have several advantages: 1) a shorter acquisition time (usually 2–3 min), 2) no need for administration of any contrast agent, and 3) available on most commercial scanners.[3] Studies have shown various utilities of the DWI sequence. DWI was proposed as a complementary adjunct sequence to improve breast MRI accuracy and to decrease unnecessary biopsies.[4] DWI with apparent diffusion coefficient (ADC) mapping was used as a quantitative imaging biomarker for prediction of immunohistochemical receptor status, proliferation rate, and molecular subtypes of breast cancer.[5] Several reported studies have shown promising effects of DWI for the detection and characterization of breast cancer.[6] In a recent study, DWI-derived radiomic signatures have been used to differentiate axillary lymph node metastasis in invasive breast cancer patients, showing a comparable performance with DCE sequences.[7]

In quantitative studies using breast MRI, segmenting the whole-breast region is an essential pre-processing step, which not only can facilitate breast tissue and/or tumor segmentation

within the breast region, but also is needed in order to compute some quantitative measures of imaging biomarkers. Many quantitative and radiomic breast MRI analyses require computing imaging features over the whole-breast region. For instance, volumetric percentage breast density (i.e., fibroglandular tissue) and background parenchyma contrast enhancement may be computed over the whole-breast region in investigating breast cancer risk biomarkers.[8] In the deep learning era, images of the segmented whole breast are highly needed in order to have deep learning models focus on effective breast regions.[9] However, the excessive effort required for manual segmentation makes it impractical. Automated whole-breast segmentation has been studied but this is a challenging problem, mainly due to the low contrast of signal intensity along breast chest wall boundaries. Several studies have developed computational methods for automated whole-breast segmentation in breast DCE MRI.[10-12] Wu et al. proposed a segmentation pipeline based on the fuzzy c-means method.[10] Gubern-Merida et al. proposed a framework to segment the breast and fibroglandular tissue simultaneously.[11] A segmentation method using sectional dynamic programming has been developed and evaluated in DCE MRI.[12] More recently, deep learning-based segmentation methods were also introduced to segment the breast and fibroglandular tissue, showing an encouraging performance.[13]

While several studies have focused on breast DCE MRI, the whole-breast segmentation in breast DWI MRI is underdeveloped. To the best of our knowledge, we have not seen fully-automated methods in the literature for such segmentation on breast DWI. Compared with DCE MRI, DWI MRI is even more challenging for whole-breast segmentation, due to the overall poorer imaging quality (blurred appearance), more severe noise/artifacts, and lower imaging resolution (Fig. 1). This is particularly true for images acquired in relatively older protocols/machines. Even if DWI and DCE image sequences are usually acquired during the same session, it is very difficult to directly translate the whole-breast segmentation masks obtained from the DCE sequences to the corresponding DWI sequences, due to the differences in imaging dimension (slice number and thickness), quality, field of view, and the difficulty of precise intersequence registration. The purpose of this study was to develop a deep/transfer learning based segmentation approach for DWI MRI scans and conduct an extensive assessment on four imaging datasets from both internal and external sources.

Materials and Methods

Datasets

We performed a retrospective study that was compliant with the Health Insurance Portability and Accountability Act (HIPAA) and received Institutional Review Board (IRB) approvals. Informed consent from patients was waived due to the retrospective nature of this study. We studied a total of 98 patients (144 MRI scans; 11,035 slices) of four different breast MRI datasets from two different institutions, where the breast DCE and DWI MRI sequences were used for developing and evaluating our segmentation methods. The detailed information for each dataset is presented in the following:

- Dataset 1: 4,251 2D DCE MRI slices from 39 breast MRI scans of 39 patients (source: University of Pittsburgh, Pittsburgh, PA). We used the pre-contrast sequences in this dataset. All scans were generated by a 1.5T MR scanner

(Avanto, Siemens, Erlangen, Germany). The parameters of the DCE sequences were: repetition time / echo time (TR/TE) = 4.74/1.39 msec, slice thickness = 1.2 mm, field of view (FOV) = 448 × 448 mm. All scans in this dataset are from patients without a breast cancer diagnosis by the time of the study. All of the slices were manually segmented by an expert radiologist (R.C., with 11-year experience of breast MRI) to outline the whole-breast region. There were no DWI sequences acquired in this dataset for historical reasons.

- Dataset 2: 431 2D DCE MRI slices selected from 20 breast MRI scans of 20 patients (source: University of Pittsburgh, Pittsburgh, PA). The imaging was acquired on a 1.5T MR scanner (Signa HDxt, GE Medical Systems, Chicago, IL). The parameters of the DCE sequences were as follows: TR/TE = 6.44/30.10 msec, slice thickness = 1.6 mm, FOV = 271 × 271 mm. No DWI MRI sequences were acquired in this dataset. To reduce the labor for manual segmentation, we selected a subset of 431 slices as a sample set; in each scan the interval of two selected slices was 10; thus, these selected subset slices are still representative of the full scans. Manual breast segmentation was performed by an expert radiologist (Y.G., with 15-year experience of breast MRI).

- Dataset 3: 6,343 2D DWI MRI slices from 75 MRI scans of 29 patients were used as the main target dataset for DWI segmentation study (source: University of Pittsburgh, Pittsburgh, PA). All these MRI scans were performed using a 1.5T MR scanner (Signa HDxt, GE Medical Systems) with the following specifications: TR/TE = 4000/77 msec, slice thickness = 5 mm, FOV = 256 × 256 mm, b1 = 150 s/mm^2, b2 = 600 s/mm^2. All scans in this dataset were from patients without breast cancer diagnosis. To reduce labor for manual segmentation, 15 scans from 15 patients were randomly selected from the full cohort to do manual breast segmentation by expert radiologists on a total of 406 slices, which are the slices that are most clinically relevant/representative for clinical diagnosis according to an expert radiologist (B.Z., with 13-year experience of breast MRI).

- Dataset 4: 10 representative 2D DWI MRI slices selected from 10 breast MRI scans from an external institution (source: China Medical University, Shenyang, China). All 10 scans were from 10 patients with a breast cancer diagnosis. All MRI scans were performed using a 3.0T MR scanner (Magnetom Verio Syngo, Siemens) with the following specification: TR/TE = 9300/76 msec, FOV = 320 × 145 mm, slices = 24, slice thickness = 4 mm, matrix size = 168 × 168, intersection gap = 0 mm, parallel imaging with sensitivity encoding factor of 2, diffusion weightings = 3, b1 = 50 s/mm^2, b2 = 400 s/mm^2, b3 = 1000 s/mm^2, diffusion directions = 3, bandwidth = 1190 Hz per pixel, echo spacing = 0.95 msec, echo planar imaging (EPI) factor = 76. The 10 clinically-relevant representative slices were selected from the single-shot spin-echo echo-planar sequence and were manually segmented by an expert radiologist (R.C., with 11-year experience of breast MRI).

Deep/Transfer Learning Models for Segmentation

We adapted the 2D UNet[14] and 2D SegNet[15] models for the breast segmentation in a slice-by-slice manner. Given an input image, both models can provide a segmentation mask with the same size of the input image. The image-to-image manner enhances the segmentation performance by utilizing features learned from different levels of the deep learning models.[16] More important, given the encouraging whole-breast segmentation performance of deep learning on the DCE MRI images, and considering that labeled DWI images are limited in reality, we leveraged the transfer learning scheme to take advantage of the good performance of deep learning models on DCE MRI images.[17-19] Thus, rather than training a UNet or SegNet model from scratch on manually segmented DWI images, we pre-trained these models on our DCE MRI data (i.e., Dataset 1), and then fine-tuned the pre-trained models with the DWI data (i.e., Dataset 3).

We implemented the UNet and the SegNet in Keras[20] and Caffe[21], respectively. The whole training process can be divided into two steps, i.e., pre-training on DCE Dataset 1 and fine-tuning on DWI Dataset 3. In pre-training, the UNet model was trained in 30 epochs with a learning rate of 2e-4, and the optimizer was Adam. It took 3 hours to complete the pre-training. The SegNet model was pre-trained in 30 epochs with a learning rate of 1e-3, and the optimizer is Stochastic Gradient Descent. It took 2.5 hours to complete the pre-training process. After pre-training, both models were fine-tuned with the same configuration as pre-training and the running time for fine-tuning was about 10 minutes. During the two training steps the batch size was set to 4. Data augmentation was performed during training. Each image was randomly rotated within the range $[-10, 10]$ degrees. The range of the random shift was $[0.9, 1.1]$ of the image length and width. The image was also zoomed in by the range of $[0.9, 1.1]$. Both models were trained on a desktop computer system with the following specifications: Intel Core i7.4790 CPU@3.60GHZ with 8 GB RAM and a Titan X Pascal Graphics Processing Unit (GPU).

Evaluation and Analysis Plans

We devised an analysis plan using these four datasets as illustrated in Fig. 2. We first implemented deep learning-based segmentation on the DCE images, which was expected to perform well according to several published studies. In this procedure, DCE Dataset 1 was mainly used to train the segmentation models, evaluated internally by three-fold cross-validation on the same dataset. Furthermore, we used the independent dataset DCE Dataset 2 for an independent evaluation on the models trained on Dataset 1. The model-building process on the DCE dataset was also the pre-training stage of the entire pipeline. Second, we fine-tuned the pre-trained models by the DWI Dataset 3 and the yielded model was internally evaluated by three-fold cross-validation on the same dataset. Likewise, we used the external DWI Dataset 4 to perform an independent evaluation to the fine-tuned segmentation model on the DWI images. For both the DCE segmentation and DWI segmentation, we performed external model evaluation to test the model's generalizability on datasets never seen before. The external dataset may have very different imaging acquisition protocols and parameters and therefore serve as a more rigorous test on the models.

Statistical Analysis

The manual segmentation of the breast performed by expert radiologists (R.C., Y.G., and B.Z.) was used as "ground truth" to quantitatively evaluate the performance of the deep/ transfer learning-based models, in terms of Dice Coefficient between the manual and model-generated segmentations. The mean value and standard deviation of the Dice Coefficients were calculated to compare segmentation results separately from the UNet and SegNet to the manual segmentation. Boxplots were used to demonstrate Dice Coefficient distributions of the two models tested on the DWI datasets.

Results

Breast Segmentation on DCE Images (i.e., Model Pre-training)

For the three-fold cross-validation of the segmentation on Dataset 1, Table 1 summarizes the segmentation performance. The average Dice Coefficient of the UNet is 0.92, substantially higher than that of the SegNet model. As demonstrated in the segmentation examples (Fig. 3), the UNet model yielded fewer false positives than the SegNet when compared with their segmentations to the manual segmentation.

For the independent evaluation on DCE Dataset 2, both models still perform well, with slight drop in performance, showing a Dice Coefficient of 0.87 for UNet and 0.80 for SegNet. As shown in the examples (Fig. 4), again, SegNet generated more false positives than UNet.

Breast Segmentation on DWI Images (i.e., Using the Fine-Tuned Model)

The models for segmenting DWI images were fine-tuned on the model pre-trained on DCE images. To highlight the benefits from transfer learning, we monitored the variations of the segmentation performance metric (i.e., the Dice Coefficient) and loss for fine-tuning vs. training from scratch. As shown in Fig. 5, the Dice Coefficient of the fine-tuned model grew faster during the training process than the training from scratch; likewise, the training loss of the fine-tuned model drops faster than the training from scratch, indicating a faster convergence. Table 2 compares the Dice Coefficient of the two models tested on DWI Dataset 3, which clearly shows that the fine-tuning outperforms the training from scratch. As shown in the segmentation examples (Fig. 6), we can see visually that the UNet model did perform better than the SegNet in the DWI image segmentation. In Dataset 3, the rest scans in addition to the 15 scans with manual annotation were also tested with both UNet and SegNet. Visual assessment by expert radiologists suggested that their segmentation results were consistent with the effects of the 15 scans. When we tested the model on the external dataset, DWI Dataset 4, both models had substantially dropped performance (UNet: from 0.85 to 0.72; SegNet: from 0.77 to 0.65) in the more difficult images (Fig. 7), where the segmentation results of the UNet model are still visually reasonable. Figure 8 shows the boxplots of Dice Coefficient distribution for the two fine-tuned models tested on both DWI Dataset 3 and DWI Dataset 4.

Discussion

Whole-breast segmentation sounds simple but it is a very challenging task and is usually the first step for quantitative breast image analysis. In this work we performed a study on applying deep/transfer learning-based segmentation methods for whole-breast segmentation in DWI MRI. Considering the clinical fact that DWI images have an overall lower imaging quality than DCE MRI, while sharing some structural/ anatomical similarity, we take advantage of the transfer learning techniques and previous success of deep learning segmentation on DCE MRI data. By implementing a transfer learning strategy on a model pre-trained on DCE MRI, even the models were fine-tuned with a limited number of DWI images, the performance of the segmentation models was encouraging both in internal and external evaluation.

Overall, our study shows that the deep learning models, UNet and SegNet, can be useful for breast segmentation in DWI images. The UNet model has shown a consistently better performance than the SegNet model in our extensive experiments. While both UNet and SegNet can roughly segment the whole-breast region, SegNet failed to distinguish the boundary information of the breast compared with UNet. There may be multiple reasons behind this observation, but we conjecture the main reason lies in the different architectures of the two models. SegNet is proposed mainly for semantic segmentation on road/indoor scenes and it is mainly studied to deal with the memory vs. accuracy trade-off in achieving good segmentation performance. When an input image is fed into the network, the encoding and decoder process may lose essential information to distinguish the breast from other regions in the images. Thus, the segmentation is likely to be scarified while reducing the trainable parameters.[15] The UNet model was originally proposed to segment biomedical images specifically for the circumstances that only a limited number of training samples are available.[14] The UNet model concatenates all features from different levels during the inference process, possibly maintaining important information for segmentation that leads to a better performance in our study. However, as regards comparison of the two models, further studies are warranted on different segmentation tasks and using different datasets.

A unique strength of our study is the utilization of multiple datasets, from internal and external sources, under different protocols and scanners. These datasets have varying characteristics and therefore are ideal to serve as challenging data to test the segmentation models. Note that in Dataset 4 the images were from cancer-developed MRIs. The nature of our method is to segment the whole breast instead of a specific type of tissue or lesion; thus, it is not sensitive to whether it is cancer-affected or negative MRIs. Still, we provided the test to our method by using the cancer-affected MRIs and, in principle, our method can be further tested by using MRIs with other masses as well. In addition, compared with studies using only a single internal dataset, this study carries substantial weight in terms of its extensive evaluation. Even though there are a smaller number of images in certain dataset (such as Dataset 4), the independent nature of the data is still very valuable in this preliminary study. However, we would like to point out that we do see a noticeable performance drop when testing the models on the external datasets. This indicates the difficulty of the segmentation task itself and the challenges to further improve the models' robustness and generalizability on unseen datasets. The overall lower imaging quality of the

breast DWI images is the main source of the challenge. Along with more advanced diffusion imaging techniques implemented in newer generations of breast MRI machines, we expect a better quality of DWI images for clinical use, and therefore improved segmentation performance can be expected in the future.

Our study has some limitations. First, due to limited and expensive time from expert radiologists, we do not have the manual segmentation for the full cohorts for some datasets. For example, we have the manual annotation on 15 patients out of the 29 patients in the DWI dataset. Yet, we are still able to show quantitative assessment and promising results of the deep/transfer learning-based approach for this challenging segmentation task. In future work, we plan to perform more extensive evaluation of our method when a larger set of manual annotations is available. Second, our approach is a 2D-based slice-by-slice segmentation. There are 3D-based volumetric segmentation methods as well, such as the 3D UNet model.[22] While we feel it is important to investigate 3D-based segmentation methods and possibly to compare with the 2D methods, the large variations on the slice number that have manual annotations prevented us from doing such an experiment at this stage of our research. In principle, we expect all breast MRI scans are segmented manually for all slices, so that across different scans we can have a reasonably comparable volumetric data to train a 3D model. Unfortunately, as stated in the Datasets subsection, there are only a few slices segmented in certain scans and the number of segmented slices vary a lot across scans. Nevertheless, we suspect that a 3D deep learning model would perform better than a 2D model because it may use spatial continuity information in the segmentation. We plan to get more manual segmentation on the datasets to be able to evaluate 3D-based models in our next steps. In addition to UNet and SegNet, other 2D or 3D deep learning-based segmentation methods may also be investigated for this specific segmentation task in future work. Finally, not all DWI images were manually segmented and selected for testing, representing a limit in validation that may be addressed in future work.

In conclusion, we investigated automated whole-breast segmentation on breast DWI data. We leveraged transfer learning of two deep learning models pre-trained on DCE MRI and used four different datasets from multiple institutions for internal and independent evaluation, showing a promising segmentation performance. This work provides a practical approach for automatic whole-breast segmentation, which sheds light on applying a deep learning-based method on whole-breast segmentation for DWI MRI scans across different MRI protocols and scanners. DWI may have the potential to augment/replace DCE sequences and we anticipate that our approach can provide a promising computerized toolkit to help enhance computer-aided quantitative analyses of breast DWI images.

Acknowledgment

References

1. Arefan D, Talebpour A, Ahmadinejhad N, et al. Automatic breast density classification using neural network. J Instrum 2015;10:T12002.

2. Radbruch A Are some agents less likely to deposit gadolinium in the brain? Magn Reson Imaging 2016;34:1351–1354. [PubMed: 27629022]

3. Partridge SC, Nissan N, Rahbar H, et al. Diffusion-weighted breast MRI: Clinical applications and emerging techniques. J Magn Reson Imaging 2017;45:337–355. [PubMed: 27690173]

4. Rahbar H, Zhang Z, Chenevert T, et al. Utility of diffusion-weighted Imaging to decrease unnecessary biopsies prompted by breast MRI: A trial of the ECOG-ACRIN cancer research group (A6702). Clin Cancer Res 2019;25:1756–1765. [PubMed: 30647080]

5. Horvat J, Bernard-Davila B, Helbich T, et al. Diffusion-weighted imaging (DWI) with apparent diffusion coefficient (ADC) mapping as a quantitative imaging biomarker for prediction of immunohistochemical receptor status, proliferation rate, and molecular subtypes of breast cancer. J Magn Reson Imaging 2019 [Epub ahead of print].

6. Cheeney S, Rahbar H, Dontchos BN, et al. Apparent diffusion coefficient values may help predict which MRI-detected high-risk breast lesions will upgrade at surgical excision. J Magn Reson Imaging 2017;46:1028–1036. [PubMed: 28181343]

7. Chai R, Ma H, Xu M, et al. Differentiating axillary lymph node metastasis in invasive breast cancer patients: A comparison of radiomic signatures from multiparametric breast MR sequences. J Magn Reson Imaging 2019 [Epub ahead of print].

8. Wu S, Berg WA, Zuley ML, et al. Breast MRI contrast enhancement kinetics of normal parenchyma correlate with presence of breast cancer. Breast Cancer Res 2016:18:76. [PubMed: 27449059]

9. Mohamed AA, Berg WA, Peng H, Luo, et al. A deep learning method for classifying mammographic breast density categories. Med Phys 2018;45:314–321. [PubMed: 29159811]

10. Wu S, Weinstein SP, Conant EF, et al. Automated fibroglandular tissue segmentation and volumetric density estimation in breast MRI using an atlas-aided fuzzy C-means method. Med Phys 2013;40:122302. [PubMed: 24320533]

11. Gubern-Merida A, Kallenberg M, Mann RM, et al. Breast segmentation and density estimation in breast MRI: A fully automatic framework. IEEE J Biomed Health Inform 2015;19:349–357. [PubMed: 25561456]

12. Jiang L, Hu X, Xiao Q, et al. Fully automated segmentation of whole breast using dynamic programming in dynamic contrast enhanced MR images. Med Phys 2017;44:2400–2414. [PubMed: 28375584]

13. Dalmis MU, Litjens G, Holland K, et al. Using deep learning to segment breast and fibroglandular tissue in MRI volumes. Med Phys 2017;44:533–546. [PubMed: 28035663]

14. Ronneberger O, Philipp F, Brox T. U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham, Switzerland: Springer; 2015.

15. Badrinarayanan V, Kendall A, Cipolla R. SegNet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE Trans Pattern Anal Mach Intell 2017;39:2481–2495. [PubMed: 28060704]

16. Wang J, Lu J, Qin G, et al. A deep learning-based auto segmentation of rectal tumors in MR images. Med Phys 2018;45:2560–2564. [PubMed: 29663417]

17. Aboutalib SS, Mohamed AA, Zuley ML,et al. Do pre-trained deep learning models improve computer-aided classification of digital mammograms? In: Medical Imaging: Computer-Aided Diagnosis of SPIE, 2018.

18. Kermany DS, Goldbaum M, Cai W, et al. Identifying medical diagnoses and treatable diseases by image-based deep learning. Cell 2018;172:1122–1131. [PubMed: 29474911]

19. Zhou Z, Shin J, Zhang L, et al. Fine-tuning convolutional neural networks for biomedical image analysis: Actively and incrementally. In: Proc IEEE CVPR 2017:7340–7351.

20. Chollet F. Keras. 2015. https://keras.io.

21. Jia Y, Shelhamer E, Donahue J, et al. Caffe: Convolutional architecture for fast feature embedding. In: Proceedings of the 22nd International Conference on Multimedia of ACM, 2014.

22. Çiçek Ö, Abdulkadir A, Lienkamp SS, et al. 3D U-Net: Learning dense volumetric segmentation from sparse annotation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham, Switzerland: Springer; 2016.
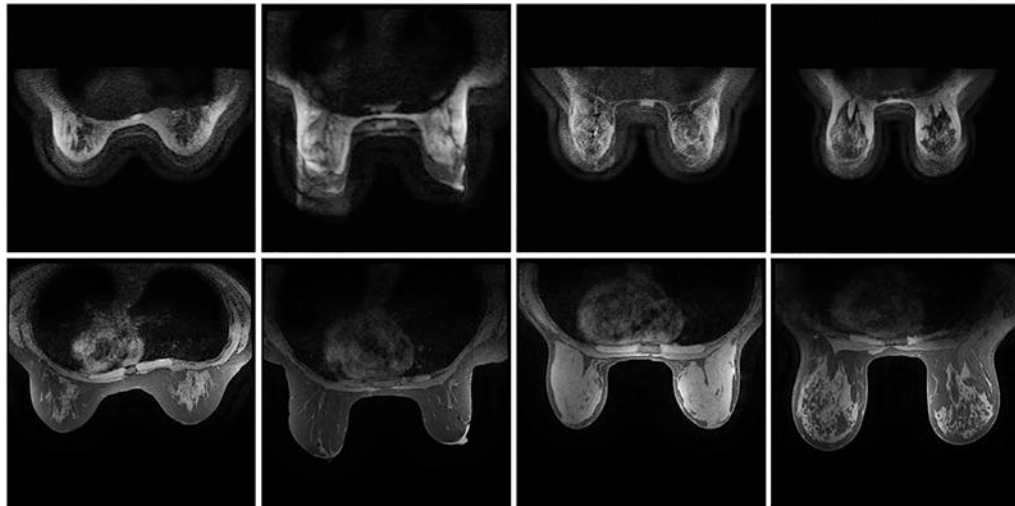
FIGURE 1:

Comparison of breast DWI (top row) and DCE (bottom row) MRI images from four different patients. Each column is a DWI center slice and a DCE center slice from the same patient. DWI images have an overall lower imaging quality (e.g., blurred appearance, more noise/artifacts, and lower resolution) and different fields of view in comparison with the DCE images, posing additional challenges for automated whole-breast segmentation in DWI images than in DCE images.
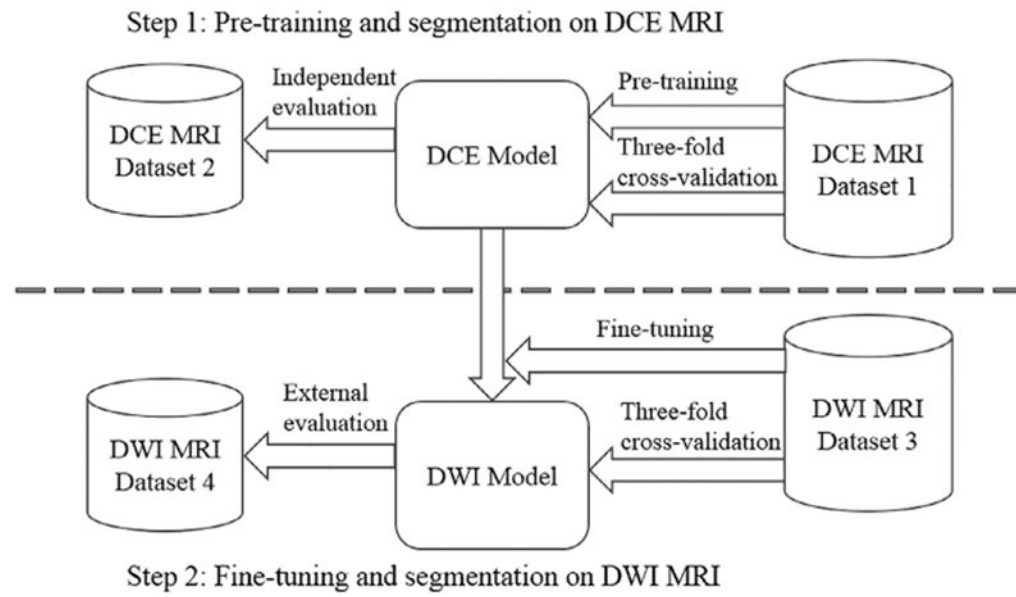
FIGURE 2:
Breast segmentation method development and evaluation pipeline.
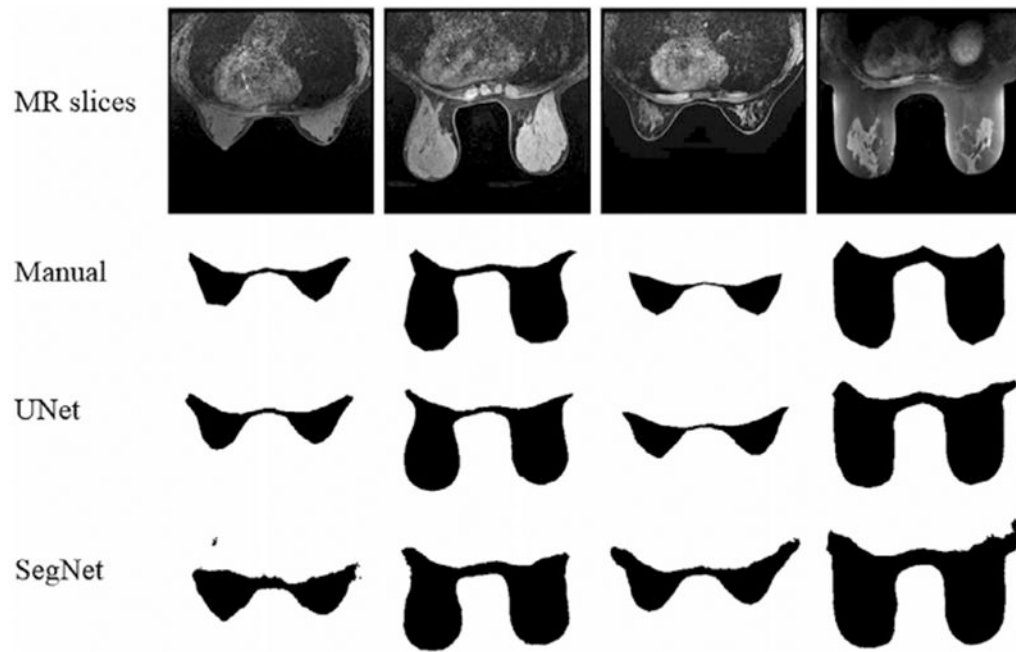
FIGURE 3:
Selected segmentation results from four different patients using UNet and SegNet on Dataset 1. As can be seen, the UNet model yielded fewer false positives than the SegNet when compared their segmentations with the manual segmentation in DCE MRI.
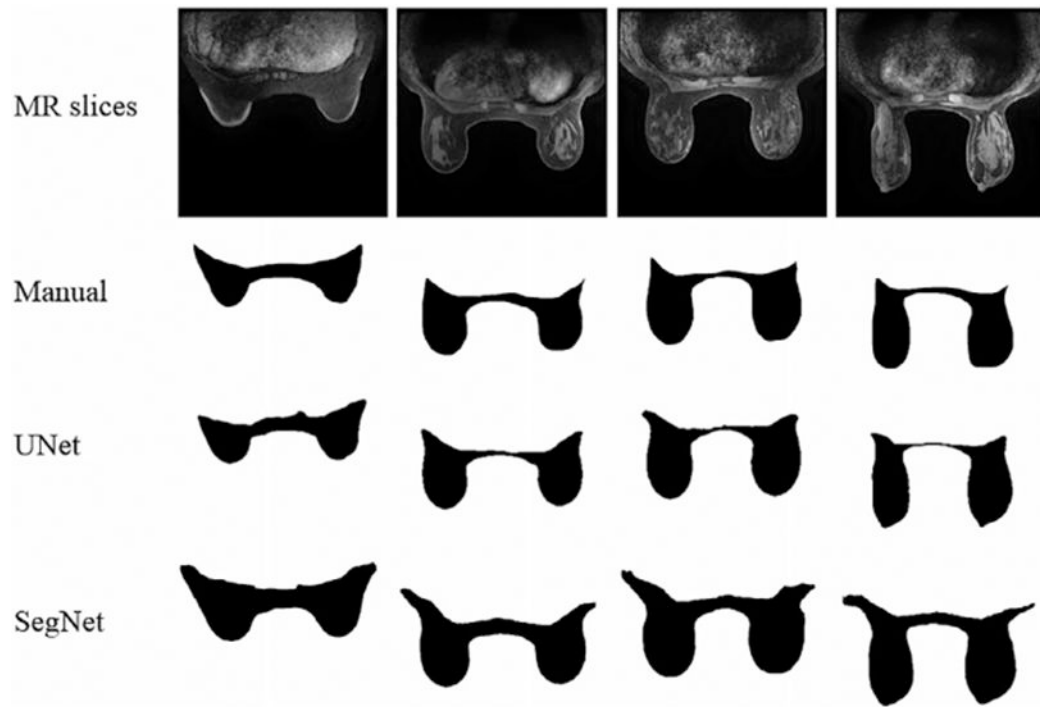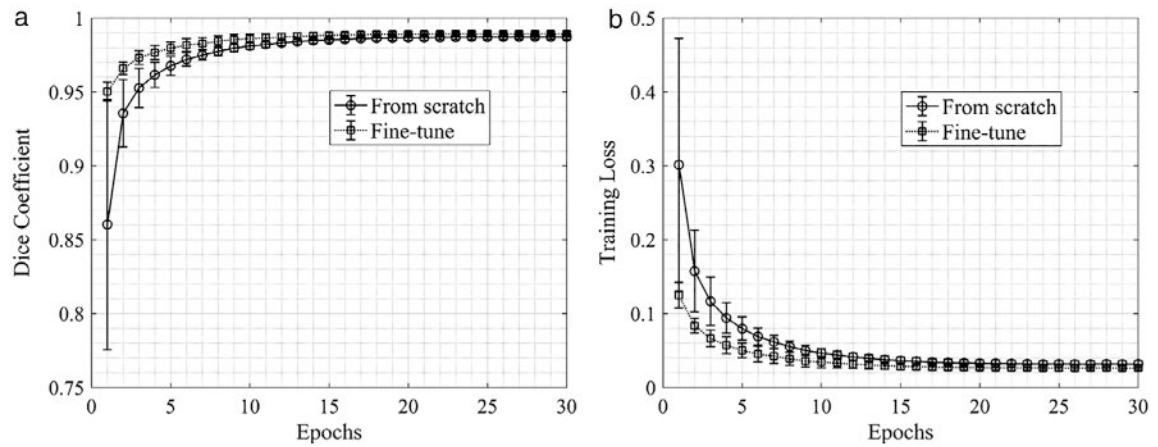
FIGURE 4:
Selected segmentation results from four different patients using UNet and SegNet on
independent never-seen Dataset 2, still showing encouraging segmentation performance.

FIGURE 5:
Dice Coefficient (a) and training loss (b) curves during training the UNet model on DWI Dataset 3 (training from scratch vs. fine-tuning).
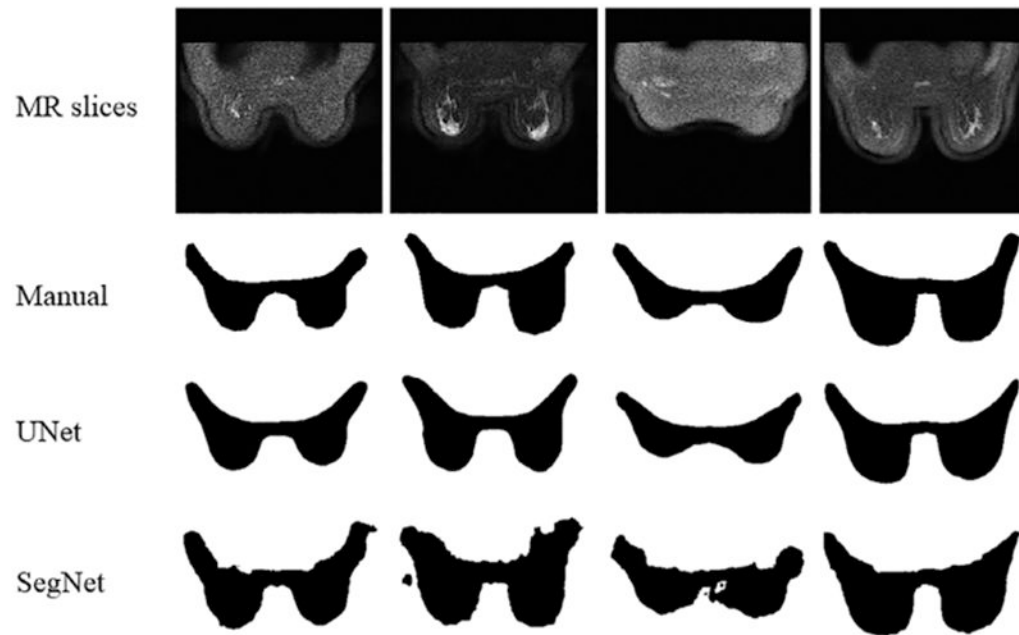
FIGURE 6:

Selected segmentation results from four different patients using UNet and SegNet on Dataset 3. In this evaluation the UNet model performs better than the SegNet in the DWI image segmentation.
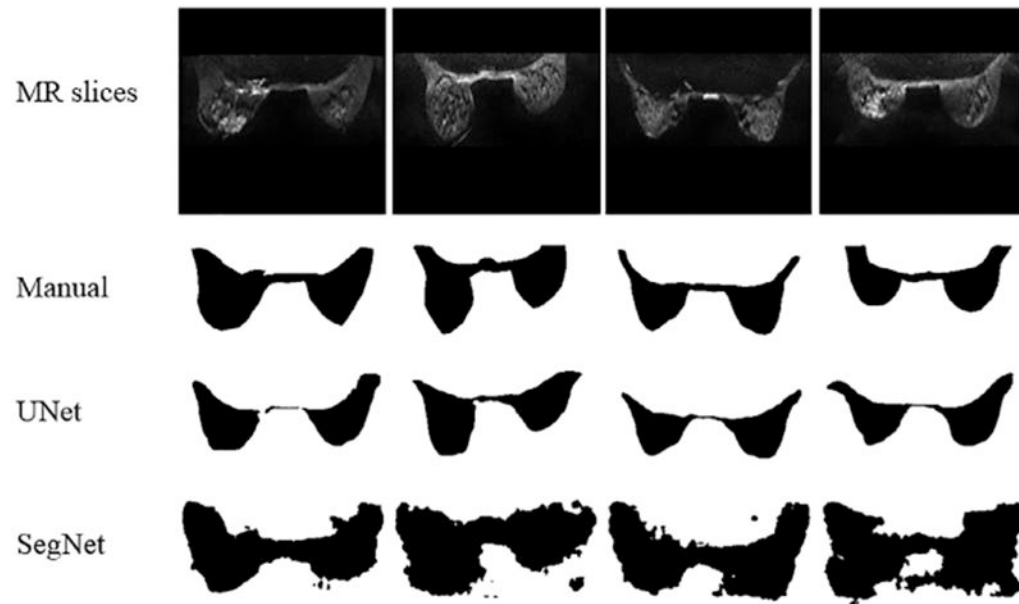
FIGURE 7:

Selected segmentation results from four different patients using UNet and SegNet on external never-seen Dataset 4. Although the image quality is poor, we can observe that the UNet and the SegNet trained by the proposed pipeline can still capture the shape of the whole-breast in DWI MRI, where the results of the UNet model are more visually reasonable.
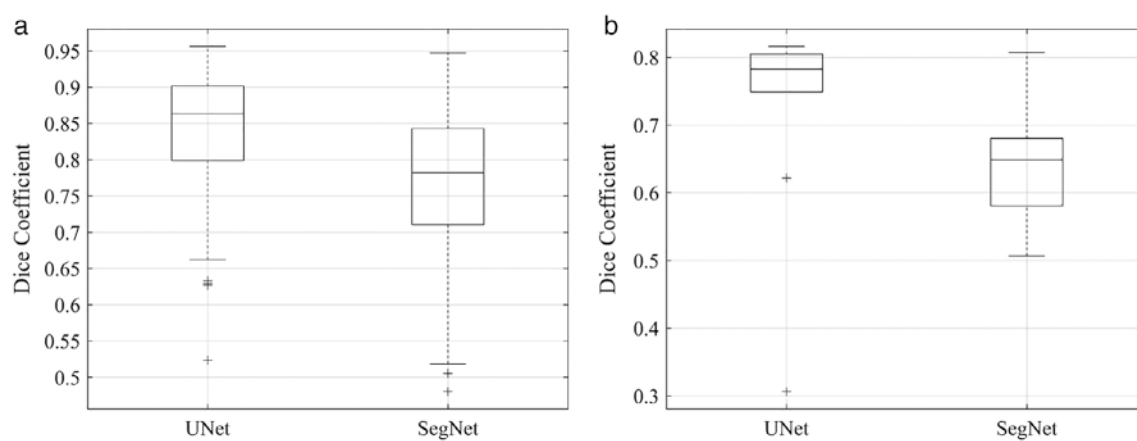
FIGURE 8:

Boxplots of the Dice Coefficient distribution for the fine-tuned UNet and SegNet evaluated on DWI Dataset 3 (a) and DWI Dataset 4 (b).

TABLE 1.

Dice Coefficients (Mean ± Standard Deviation) of the UNet and SegNet on DCE MRI Datasets

Data	Model	Dice Coefficient
DCE dataset 1		
	UNet	0.92 ± 0.07
	SegNet	0.84 ± 0.11
DCE dataset 2		
	UNet	0.87 ± 0.06
	SegNet	0.80 ± 0.06

TABLE 2.

Dice Coefficients (Mean ± Standard Deviation) of UNet and SegNet on DWI MRI Datasets

Data	Model	Training method	Dice Coefficient
DWI dataset 3			
	UNet	From scratch	0.69 ± 0.12
	SegNet	From scratch	0.73 ± 0.18
	UNet	Fine-tune	0.85 ± 0.07
	SegNet	Fine-tune	0.77 ± 0.09
DWI dataset 4			
	UNet	Fine-tune	0.72 ± 0.16
	SegNet	Fine-tune	0.65 ± 0.10