

Fat-suppressed breast MRI synthesis for domain adaptation in tumour segmentation

Lidia Garrucho^{†a,b,*}, Eve Deleuge^{†a,c}, Richard Osuala^{a,d,e}, Dimitri Kessler^a, Kaisar Kushibar^a, Oliver Diaz^a, Karim Lekadir^a, Laura Igual^{a,f}

^a*Artificial Intelligence in Medicine Lab (BCN-AIM), Facultat de Matemàtiques i Informàtica, Universitat de Barcelona, Barcelona, Spain*

^b*Department of Oncology-Pathology, Karolinska Institutet, Stockholm, Sweden*

^c*Ecole Normale Supérieure de Paris Saclay, Gif-Sur-Yvette, France*

^d*Helmholtz Center Munich, Munich, Germany*

^e*Technical University of Munich, Munich, Germany*

^f*Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain*

ARTICLE INFO

Keywords: Synthetic data, breast cancer, deep learning, generative models, image-to-image translation, fat suppression

ABSTRACT

Heterogeneity in dynamic contrast-enhanced breast MRI acquisition protocols hinders the generalization of automatic tumour segmentation tools. While fat-suppressed MRI acquisition is common, some vendors do not provide these sequences, making a segmentation model trained with fat-suppressed images unusable for non-fat-suppressed cases. In this study, we propose two strategies to alleviate this issue. The first approach involves translating non-fat-suppressed to fat-suppressed breast MRI. The second approach integrates synthetic non-fat-suppressed MRI into the training pipeline of tumour segmentation models. Our experimental results demonstrate that both approaches significantly improve segmentation performance on non-fat-suppressed MRI, suggesting that domain adaptation techniques based on image synthesis can enhance the accuracy and reliability of tumour segmentation in breast MRI.

1. Introduction

A frequent issue in breast MRI encompasses the high signal from fat, which can obscure cancerous lesions thereby impacting accurate diagnosis and timely treatment. To address this, fat suppression techniques such as fat-saturation and subtraction, are commonly used to reduce artifacts and enhance gadolinium contrast visualization, crucial for determining malignancy Delfaut et al. (1999). The application of fat suppression techniques varies widely among hospitals, patients, and scans due to factors such as clinical context, diagnostic goals, breast fat content, magnetic field, technology, and institutional protocols Kuhl (2007); Mori et al. (2020). This variability can degrade the performance of automated medical image analysis methods, particularly deep learning models, which are sensitive to domain shifts. Therefore, research into consistent automated

fat-suppression methodologies for clinical application is essential as a complement or even an alternative to existing methods.

Volumetric tumour segmentation is crucial for assessing cancer volume and extension. Currently, deep learning-based segmentation methods, such as *nnU-Net* Isensee et al. (2021), have shown excellent results in various biomedical image segmentation tasks. However, these methods are data-intensive and rely on the availability of ground-truth segmentations for training. In T1-weighted breast MRI, publicly available ground-truth segmentations of breast cancer are predominantly restricted to images with fat suppression. However, as mentioned above, imaging protocols differ across hospitals, complicating the generalization of these models to non-fat-suppressed T1-weighted images, as well as across different fat-suppressed breast MRI domains.

As depicted in Fig. 1, given automatic tumour segmentation models trained on fat-suppressed MRI, we propose two domain adaptation (DA) techniques to improve the performance of tu-

*†Both authors contributed equally to this work.

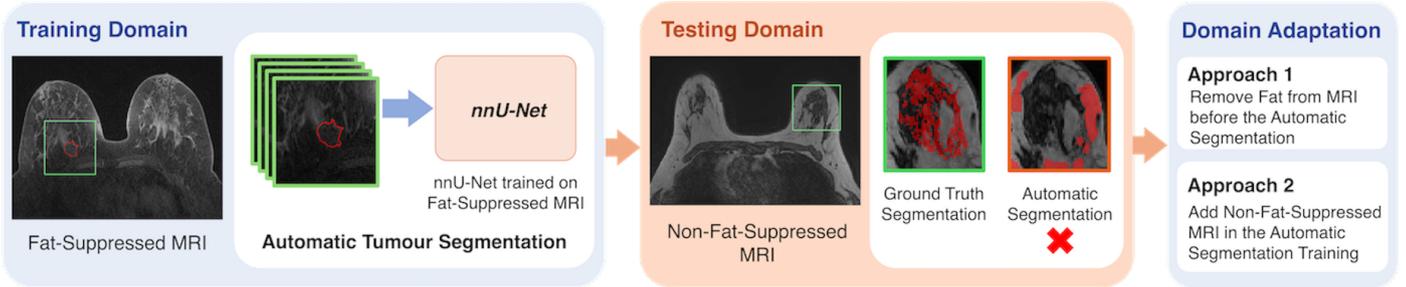


Fig. 1: Overview of our proposed domain adaptation pipeline illustrating two alternative methods validated in this study for volumetric tumour segmentation in non-fat-suppressed T1-weighted breast MRI.

tumour segmentation on non-fat-suppressed MRI. The first approach involves applying fat suppression to non-fat-suppressed MRI using Generative Adversarial Networks (GANs) Goodfellow et al. (2014); Li et al. (2022); Zhang et al. (2024); Han et al. (2024) and subsequently performing tumour segmentation on the synthetic fat-suppressed images. This largely unexplored approach is motivated by some promising initial results Mori et al. (2020) reported for synthetic fat-suppressed images, which, however, have not been assessed in terms of their utility for clinical applications. To this end, our goal is to evaluate the effectiveness of synthetic fat-suppressed MRI for tumour segmentation, potentially as useful replacement for real non-fat-suppressed MRI. Our second approach constitutes the first study to include and evaluate synthetic non-fat-suppressed MRI in automatic tumour segmentation model training pipelines. In this case, instead of performing fat suppression, the GAN learns to inpaint fat into the real fat-suppressed MRI. Incorporating synthetic non-fat-suppressed images into the training pipeline can eliminate the need for MRI post-processing steps (synthesizing fat-suppressed MRI from non-fat-suppressed ones), thereby reducing computational resources and time. In sum, the key contributions of our work are as follows:

- Design, implementation and multi-metric validation of a conditional Generative Adversarial Network for synthesizing fat-suppressed and non-fat-suppressed breast MRI.
- Evaluation of the effect of domain shift in fat-suppressed T1-weighted breast MRI on tumour segmentation.
- Contribute with the first in-depth comparative analysis and validation of synthetic non-fat-suppressed MRI (i) as target inference domain and (ii) as training data augmentation method, demonstrating its potential to improve tumour segmentation performance.

2. Materials and Methods

2.1. Datasets

The MAMA-MIA dataset Garrucho et al. (2024), a collection of 1506 ground-truth breast MRI tumour segmentations from four public datasets Saha et al. (2021) Newitt and Hylton (2016a) Li (2022) Newitt and Hylton (2016b), was used in this study to train the automatic tumour segmentation models. The dataset encompasses 1271 axial and 235 sagittal T1-weighted dynamic contrast-enhanced scans (DCE-MRI) with fat suppression from three different vendors. The Advanced MRI Breast Lesions (AMBL) dataset Daniels et al. (2024) was used to train the GANs. AMBL contains 632 cases acquired on a 1.5T MR system between 2018 and 2021. The patients were screened with various MRI modalities and ground-truth tumour segmentations are available for 99 cases. In this study, we use the non-fat-suppressed T1-weighted MRI and the first phase (pre-contrast) of the axial T1-weighted DCE-MRI as image pairs for image synthesis. Additionally, 47 test cases with ground-truth tumour segmentations were randomly excluded from image-to-image translation training to evaluate the tumour segmentation performance in an external test set.

2.2. Fat-Suppressed and Non-Fat-Suppressed MRI Synthesis

We adopt *pix2pix* Isola et al. (2017), a conditional Generative Adversarial Network (cGAN) framework for paired image-to-image translation, was trained to synthesized fat-suppressed and non-fat-suppressed 2D axial MRI slices. As depicted in Fig. 2, we design a pre-processing pipeline involving several steps to assemble MRI image pairs for our training setup. First, the pre-contrast fat-suppressed T1-weighted DCE-MRI phase and the pre-contrast non-fat-suppressed T1-weighted MRI were registered using rigid-affine registration Ourselin et al. (2001). Next, the MRI were resampled to isotropic pixel spacing to ensure consistent resolution across all three axes (axial, sagittal,

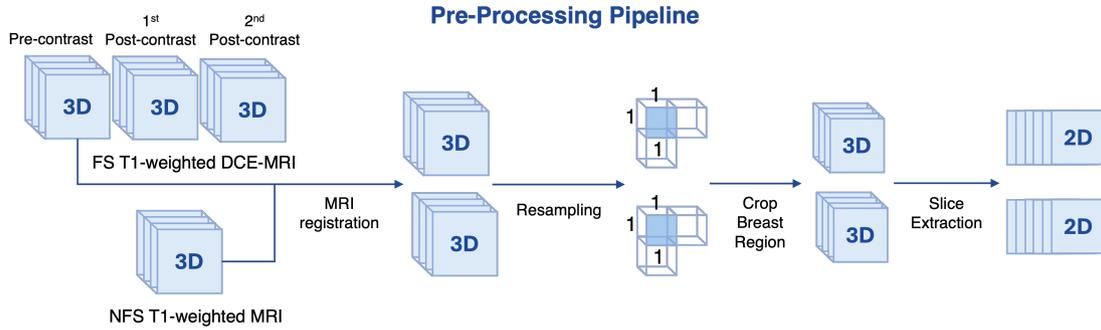


Fig. 2: Steps to extract the paired fat-suppressed (FS) and non-fat-suppressed (NFS) MRI slices to train our conditional GAN models.

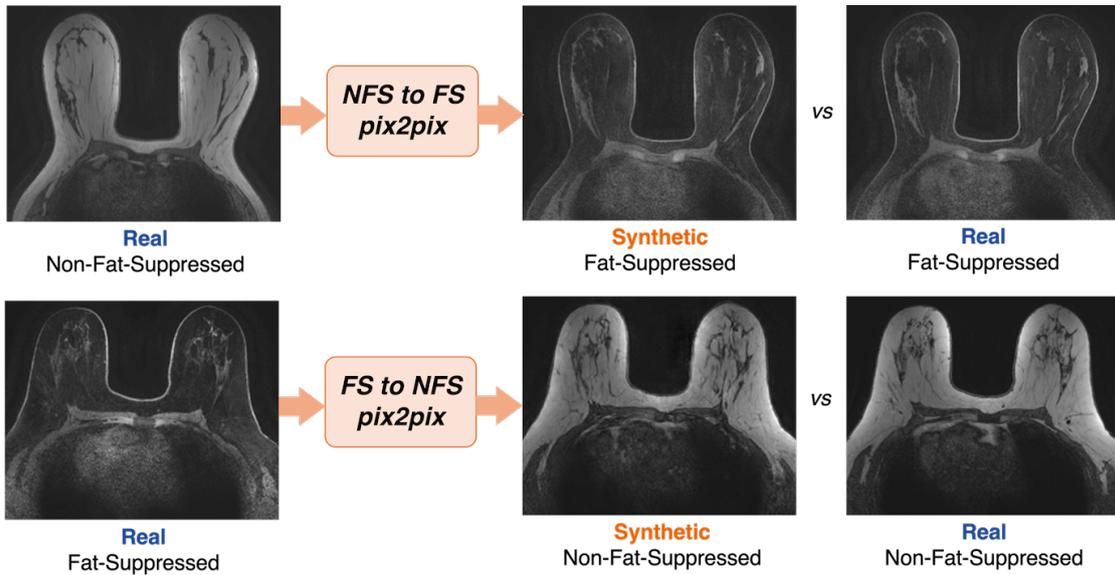


Fig. 3: Bi-directional image-to-image translation using *pix2pix* for synthesizing either FS or NFS T1-weighted MRI slices, illustrated alongside their real counterparts.

and coronal), utilizing B-Spline interpolation implemented using the SimpleITK library Lowekamp *et al.* (2013). Lastly, the MRI were cropped to the breast region using Otsu’s thresholding to discard the background before extracting slices and pairing corresponding axial MRI slices.

After image pre-processing, the AMBL dataset was split in 471 MRI cases for training and 160 for testing. Two *pix2pix* models were trained using paired fat-suppressed and non-fat-suppressed MRI 2D slices with input size 256x256, as exemplified by Fig. 3. We follow the *pix2pix* hyperparameter setup Isola *et al.* (2017), training during 200 epochs, using serial batches, and the Adam optimizer with $\beta_1 = 0.5$. The training schedule consisted of 100 epochs with an initial learning rate of 0.0002, followed by 100 epochs with a linearly decreasing learning rate.

2.3. Automatic Tumour Segmentation

We integrate *nnU-Net* Isensee *et al.* (2021) as an automatic tumour segmentation model into our pipeline. Tumour volumes

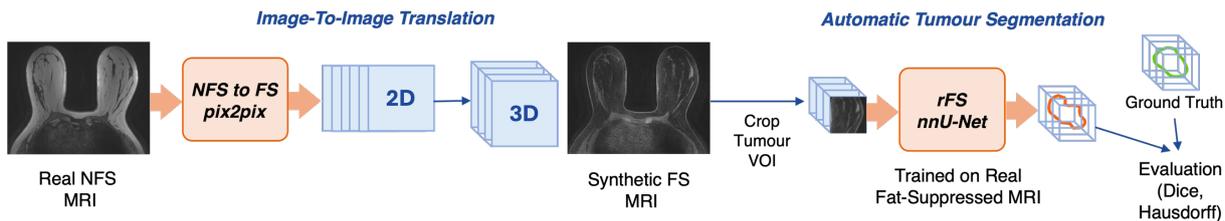
(VOI) of 1506 real fat-suppressed pre-contrast T1-weighted MRI with an additional 20% margin were used to train a 3D *nnU-Net* in a 5-fold cross-validation setting. The training parameters include using z-score normalization, isotropic pixel spacing, an initial learning rate of 1e-2, and a weight decay of 3e-5 during 1000 epochs.

As shown in Fig. 4, we design two different DA approaches to improve the automatic tumour segmentation in out-of-domain non-fat-suppressed MRI.

2.3.1. Approach 1: Tumour Segmentation on Synthetic Fat-Suppressed MRI

The first approach involves performing image-to-image translation prior to automatic tumour segmentation. The *pix2pix*, trained to translate non-fat-suppressed to fat-suppressed MRI (*NFS to FS pix2pix*), is applied to synthesize 2D axial slices that are stacked to respective 3D MRI volumes back to their original space. Further, these synthetic MRI vol-

Approach 1: Automatic Tumour Segmentation on Synthetic Fat-Suppressed MRI



Approach 2: Automatic Tumour Segmentation using Synthetic Non-Fat-Suppressed MRI as Data Augmentation

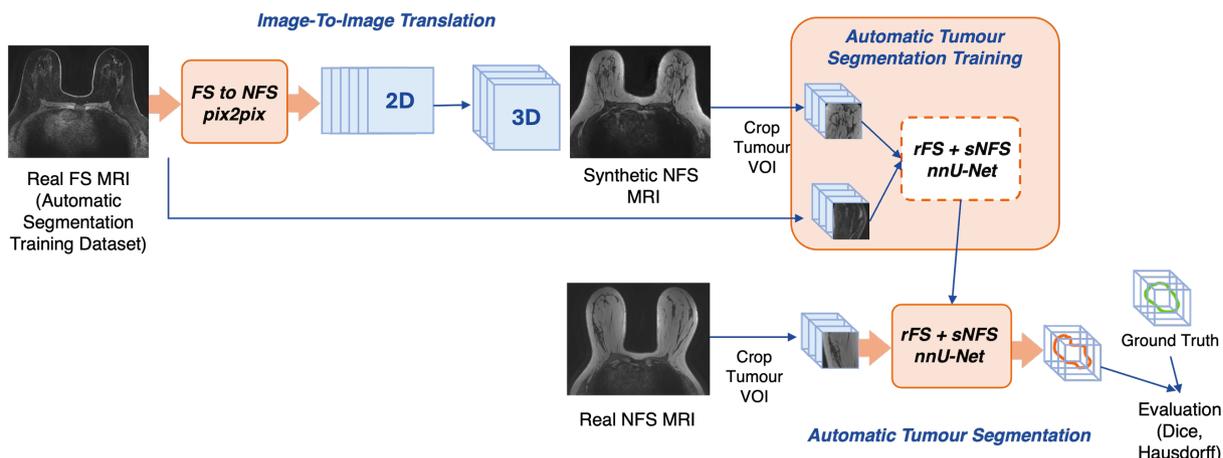


Fig. 4: Our domain adaptation approaches for tumour segmentation of non-fat-suppressed (NFS) T1-weighted breast MRI.

umes are then cropped to extract the volume of interest to segment using the *rFS nnU-Net*.

2.3.2. Approach 2: Synthetic Non-Fat-Suppressed MRI Data Augmentation

The second approach involves the addition of synthetic non-fat-suppressed MRIs in the *nnU-Net* training pipeline. Synthetic non-fat-suppressed MRI are generated using the *FS to NFS pix2pix* for each of the 1271 axial MRI cases in the MAMA-MIA dataset. The final segmentation model (*rFS+sNFS nnU-Net*) is, thus, trained using real fat-suppressed (*rFS*) and synthetic non-fat-suppressed (*sNFS*) MRI.

2.4. Evaluation Metrics and Statistical Analysis

The performance of the *pix2pix* models is evaluated using Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM) Wang et al. (2004), Fréchet Inception Distance (FID) Heusel et al. (2017), and Fréchet Radiomics Distance (FRD) Osuala et al. (2024). High PSNR values indicate close resemblance to ground-truth images, while SSIM aligns with human perception, crucial for clinical evaluation. FID assesses

image realism, and FRD evaluates dataset similarity based on radiomics features, providing a more accurate complementary assessment for breast MRI. These metrics offer insights into pixel-level similarity, structural fidelity, realism, and radiomics feature alignment.

Automatic tumour segmentation performance is evaluated using the Dice Coefficient and the 95th percentile of the Hausdorff Distance (HD95). The Dice Coefficient measures the overlap between predicted and ground-truth segmentations, indicating accuracy, while HD95 assesses boundary accuracy by measuring the largest distance between predicted and ground-truth boundaries, excluding outliers Huttenlocher et al. (1993).

Statistical differences in segmentation performances are assessed using the Wilcoxon signed-rank test with Bonferroni Correction, a non-parametric test suitable for non-normally distributed data common in biomedical research Wilcoxon (1992).

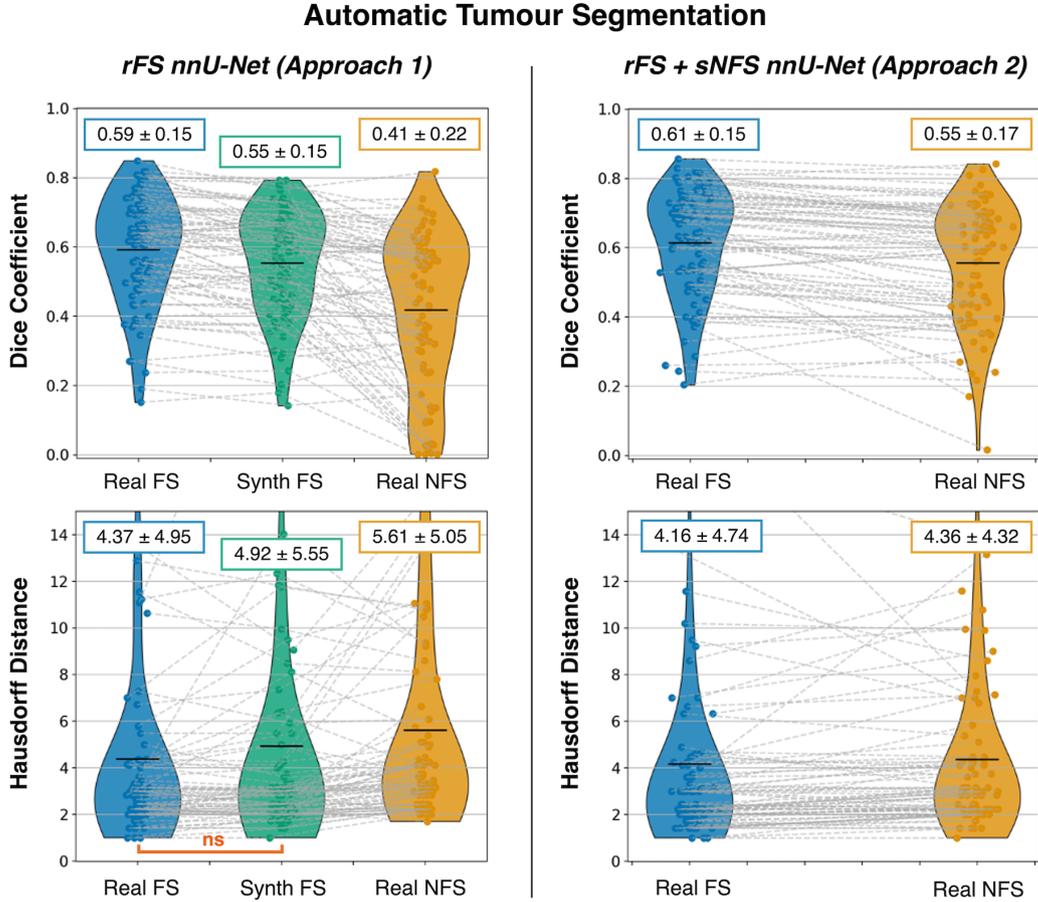


Fig. 5: Comparison of automatic tumour segmentation evaluation metrics for both DA approaches on the test set. 'FS' stands for Fat-Suppressed, 'NFS' for Non-Fat-Suppressed, and 'Synth' for Synthetic. *ns* indicates no statistical difference between metrics.

Table 1: Performance metrics for *pix2pix* models synthesizing fat-suppressed (FS) and non-fat-suppressed (NFS) MRI.

<i>pix2pix</i>	PSNR ↑	SSIM ↑	FID ↓	FRD ↓
<i>NFS to FS</i>	25.58 ± 3.50	0.75 ± 0.11	5.54	13.24
<i>FS to NFS</i>	26.39 ± 0.46	0.69 ± 0.26	34.10	52.14

3. Experiments and Results

3.1. Fat-Suppressed and Non-Fat-Suppressed MRI Synthesis

The quantitative evaluation of both *pix2pix* models is depicted in Table 1. The *NFS to FS* model achieved a promising PSNR of 25.58 ± 3.50 , an SSIM of 0.75 ± 0.11 , an FID of 5.54, and an FRD of 13.24, indicating high pixel-level similarity, structural fidelity, and radiomic imaging biomarker accuracy. The *FS to NFS* model had a higher PSNR of 26.39 ± 0.46 but lower structural fidelity with an SSIM of 0.69 ± 0.26 . As indicated by an FID of 34.10 and an FRD of 52.14, the model generated synthetic images of high quality, however, they re-

sulted overall less realistic than their synthetic *NFS to FS* counterparts.

3.2. Automatic Tumour Segmentation

The mean 5-fold cross-validation Dice coefficients for the automatic tumour segmentation models trained on MAMA-MIA were 0.76 ± 0.01 for *rFS nnU-Net* in Approach 1 (A1) and 0.76 ± 0.01 for *rFS + sNFS nnU-Net* in Approach 2 (A2), indicating that incorporating synthetic non-fat-suppressed (Synth NFS) MRI did not negatively affect overall performance. The segmentation performance on real FS tumours in the external test set (47 cases from AMBL dataset containing 79 tumours) decreased due to domain shift (Dice of 0.59 for *rFS nnU-Net* in A1 and 0.61 for *rFS + sNFS nnU-Net* in A2).

Both DA approaches, with results summarized in Fig. 5, improved segmentation metrics compared to real non-fat-suppressed (Real NFS) images segmented with *rFS nnU-Net* in A1. Metrics show similar performance between Synth FS in A1 and Real NFS in A2 (Dice: 0.55 ± 0.15 vs. 0.55 ± 0.17 , HD: 4.92 ± 5.55 vs. 4.36 ± 4.32). The Dice coefficient in Real FS is

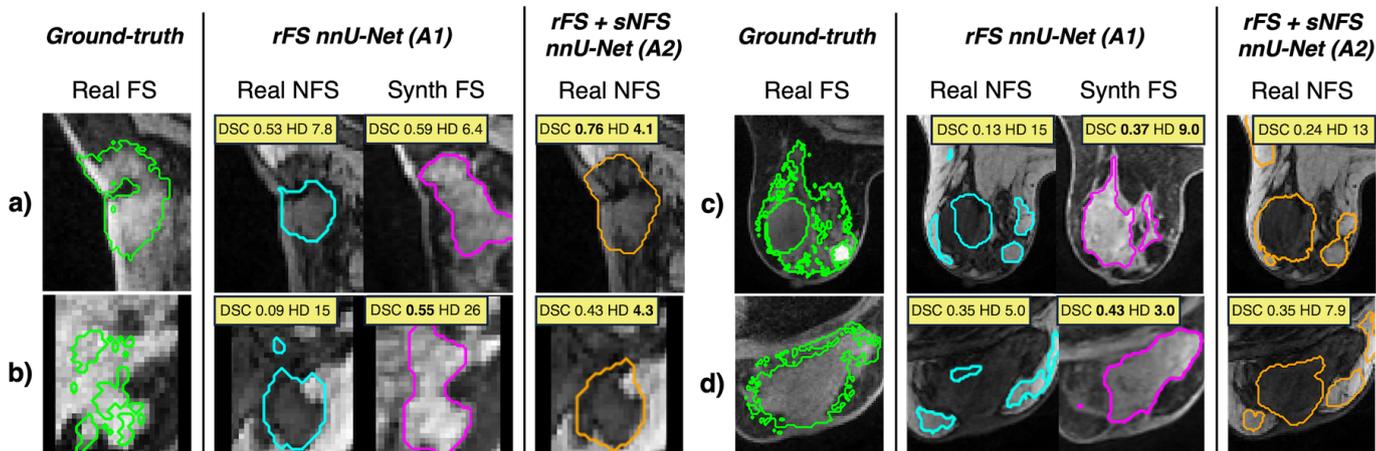


Fig. 6: Four examples (a), b), c) and d)) of tumour segmentation contours highlighting the differences between the ground-truth in real FS and the results of the DA approaches. 'A1' stands for Approach 1, 'A2' for Approach 2, 'FS' for Fat-Suppressed, 'NFS' for Non-Fat-Suppressed, 'Synth' for Synthetic, 'DSC' for Dice Coefficient, and 'HD' for Hausdorff Distance.

slightly better in A2 (0.61 ± 0.15 vs. A1: 0.59 ± 0.15), indicating that Synth NFS data augmentation improved the segmentation model. Despite statistical differences in Dice coefficient between Real FS and Synth FS, this difference is not observed in the Hausdorff Distance metric, which evaluates boundary accuracy. Fig. 6 presents qualitative results for four test cases, supporting our findings. Although the tumour segmentations are volumetric, only the middle slice of each tumour is displayed for clarity.

4. Discussion and Conclusion

Our study investigates DA techniques leveraging image synthesis to enhance tumour segmentation in non-fat-suppressed (NFS) breast MRI. We employed two distinct DA approaches: Approach 1 (A1), which runs tumour segmentation in synthetic fat-suppressed (Synth FS) MRI obtained using the *NFS to FS pix2pix* model, and Approach 2 (A2), which incorporates synthetic non-fat-suppressed (Synth NFS) MRI generated with the *FS to NFS pix2pix* model into the automatic tumour segmentation pipeline (*rFS + sNFS nnU-Net* model) to improve the tumour segmentation in real NFS MRI.

Our FS MRI synthesis with *NFS to FS pix2pix* model showed superior performance across multiple evaluation metrics. The *FS to NFS pix2pix* model, while delivering higher PSNR, faced challenges related to structural and radiomic fidelity, evident from higher FID and FRD values. Nevertheless, integrating Synth NFS MRI into the training pipeline in A2 did not compromise overall tumour segmentation performance. On the contrary, it bolstered the Dice coefficient for Real NFS, underscoring the efficacy of data augmentation with synthetic images.

Both DA approaches exhibited notable improvements in tumour segmentation metrics compared to real NFS MRI segmented with *rFS nnU-Net* in A1. Although statistical differences in Dice coefficients were observed between Real FS and Synth FS, there was no statistical difference for the Hausdorff distance, demonstrating remarkable synthetic data utility.

Our findings highlight the potential of both DA approaches for enhancing tumour segmentation on NFS MRI. Moreover, the synthetic NFS images generated by the *NFS to FS pix2pix* model offer a promising alternative to conventional fat suppression techniques. Future research aims to address current study limitations by exploring their utility in diverse clinical tasks related to breast cancer using MRI, such as comprehensive cancer detection in high-risk screening populations and prediction of treatment response. Additionally, future efforts will focus on synthesizing 3D MRI to maintain spatial continuity across 2D slices, integrating perceptual loss in the training pipeline to enhance synthetic MRI fidelity, and validating models externally with additional datasets.

5. Acknowledgments

This project was funded by the EU's Horizon 2020 programmes under grants No 952103 (EUCanImage) and No 101057699 (RadioVal). It was partially supported by Spain's Ministry of Science and Innovation (FUTURE-ES PID2021-126724OB-I00), MICINN Grant PID2022-136436NB-I00, and AGAUR Grant 2021-SGR-01104. Co-author K.K. holds a Juan de la Cierva fellowship (FJC2021-047659-I).

6. Disclosure of Interests

The authors declare no competing interests.

References

- Daniels, D., Last, D., Cohen, K., Mardor, Y., Sklair-Levy, M., 2024. Standard and delayed contrast-enhanced MRI of malignant and benign breast lesions with histological and clinical supporting data (Advanced-MRI-Breast-Lesions) (version 2) [data set]. The Cancer Imaging Archive doi:10.7937/C7X1-YN57.
- Delfaut, E.M., Beltran, J., Johnson, G., Rousseau, J., Marchandise, X., Cotten, A., 1999. Fat suppression in MR imaging: techniques and pitfalls. *Radiographics* 19, 373–382. doi:10.1148/radiographics.19.2.g99mr03373.
- Garrucho, L., Reidel, C.A., Kushibar, K., et al., 2024. MAMA-MIA: A large-scale multi-center breast cancer DCE-MRI benchmark dataset with expert segmentations. arXiv preprint URL: <https://arxiv.org/abs/2406.13844>, doi:10.48550/arXiv.2406.13844.
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2014. Generative adversarial nets, in: Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N., Weinberger, K.Q. (Eds.), *Advances in Neural Information Processing Systems*, Curran Associates, Inc., pp. 2672–2680.
- Han, L., Tan, T., Zhang, T., Huang, Y., Wang, X., Gao, Y., Teuwen, J., Mann, R., 2024. Synthesis-based imaging-differentiation representation learning for multi-sequence 3D/4D MRI. *Medical Image Analysis* 92, 103044.
- Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., Hochreiter, S., 2017. GANs trained by a two time-scale update rule converge to a local Nash equilibrium. *Advances in Neural Information Processing Systems* 30.
- Huttenlocher, D.P., Klanderman, G.A., Rucklidge, W.J., 1993. Comparing images using the Hausdorff distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 15, 850–863. doi:10.1109/34.232073.
- Isensee, F., Jaeger, P.F., Kohl, S.A.A., Petersen, J., Maier-Hein, K.H., 2021. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature Methods* 18, 203–211. doi:10.1038/s41592-020-01008-z.
- Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A., 2017. Image-to-image translation with conditional adversarial networks, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1125–1134.
- Kuhl, C., 2007. The current status of breast MR imaging part I. Choice of technique, image interpretation, diagnostic accuracy, and transfer to clinical practice. *Radiology* 244, 356–378. doi:10.1148/radiol.2442051620.
- Li, W., 2022. I-SPY 2 breast dynamic contrast enhanced MRI trial (version 1) [data set]. The Cancer Imaging Archive doi:10.7937/TCIA.D8Z0-9T85.
- Li, W., Xiao, H., Li, T., et al., 2022. Virtual contrast-enhanced magnetic resonance images synthesis for patients with nasopharyngeal carcinoma using multimodality-guided synergistic neural network. *IJROBP* 112, 1033–1044.
- Lowekamp, B.C., Chen, D.T., Ibáñez, L., Blezek, D., 2013. The design of SimpleITK. *Frontiers in Neuroinformatics* 7, 45.
- Mori, M., Fujioka, T., Katsuta, L., Kikuchi, Y., Oda, G., Nakagawa, T., Kitazume, Y., Kubota, K., Tateishi, U., 2020. Feasibility of new fat suppression for breast MRI using pix2pix. *Japanese Journal of Radiology* 38, 1075–1081. doi:10.1007/s11604-020-01012-5.
- Newitt, D., Hylton, N., 2016a. Multi-center breast DCE-MRI data and segmentations from patients in the I-SPY 1/ACRIN 6657 trials. The Cancer Imaging Archive doi:10.7937/K9/TCIA.2016.HdHpgJLK.
- Newitt, D., Hylton, N., 2016b. Single site breast DCE-MRI data and segmentations from patients undergoing neoadjuvant chemotherapy (version 3) [data set]. The Cancer Imaging Archive doi:10.7937/K9/TCIA.2016.QHsyhJKy.
- Osuala, R., Lang, D., Verma, P., Joshi, S., Tsirikoglou, A., Skorupko, G., Kushibar, K., Garrucho, L., Pinaya, W.H.L., Diaz, O., et al., 2024. Towards learning contrast kinetics with multi-condition latent diffusion models. arXiv preprint arXiv:2403.13890.
- Ourselin, S., Roche, A., Subsol, G., Pennec, X., Ayache, N., 2001. Reconstructing a 3D structure from serial histological sections. *Image and Vision Computing* 19, 25–31.
- Saha, A., Harowicz, M.R., Grimm, L.J., Weng, J., Cain, E.H., Kim, C.E., Ghate, S.V., Walsh, R., Mazurowski, M.A., 2021. Dynamic contrast-enhanced magnetic resonance images of breast cancer patients with tumor locations [data set]. The Cancer Imaging Archive doi:10.7937/TCIA.e3sv-re93.
- Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P., 2004. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing* 13, 600–612.
- Wilcoxon, F., 1992. *Individual comparisons by ranking methods*. Springer. doi:10.1007/978-1-4612-4380-9_16.
- Zhang, T., Tan, T., Han, L., et al., 2024. IMPORTANT-Net: Integrated MRI multi-parametric increment fusion generator with attention network for synthesizing absent data. *Information Fusion* 108, 102381.