Robust multi-vendor breast region segmentation using deep learning

Koen Dercksen[a], **Michiel Kallenberg**[b], **Jaap Kroes**[b]
[a]Radboud University, iCIS, Data Science, Toernooiveld 200, Nijmegen, The Netherlands, 6525 EC
[b]Screenpoint Medical BV, Toernooiveld 300, Nijmegen, The Netherlands, 6525 EC

Abstract. Semantic segmentation of breast images is typically performed as a preprocessing step for breast cancer detection by Computer Aided Diagnosis (CAD) systems. While most literature on region segmentation is based on conventional techniques like line estimation, thresholding and atlas-based approaches, such methods may have problems with generalisation. This paper investigates a robust multi-vendor breast region segmentation system for full field digital mammograms (FFDM) and digital breast tomography (DBT) using a U-Net neural network. Additionally, the effect of adding attention gates to the U-Net architecture was analysed. The proposed networks were trained and tested in a cross-validation setting on in-house FFDM/DBT data and the public INbreast datasets, comprising over 10,000 FFDM and 3,500 DBT images from five different vendors. Dice scores were obtained in the range 0.978 — 0.985, with slightly higher scores for the architecture that includes attention gates.

Keywords: digital breast imaging, deep learning, semantic segmentation.

1 Introduction

Segmentation of the breast is often the first step in computerized analysis of mammographic images. Most literature on segmentation of breast images is based on conventional techniques like line estimation, thresholding and atlas-based approaches.[1] As these methods are based on built-in logic and assumptions they may have problems with generalisation. This becomes more pronounced in the multi-vendor context, as images can vary wildly between scanners. In this work, we investigate the use of deep learning for region segmentation in full field digital mammography (FFDM) and digital breast tomography (DBT). We suspect that a deep learning approach will be able to robustly deal with multi-vendor region segmentation of background, breast tissue and pectoral muscle. Deep learning approaches are not as widespread for the task at hand. Neural networks are mostly popular in the related field of breast lesion detection, classification and/or segmentation but not so much in breast region segmentation. Rodriguez et al.[2] used a U-Net network to segment FFDM and DBT images into background, breast and pectoral muscle, showing generalisation between the two modalities. Rampun et al.[3] used a modified version of the holistic edge detection network (HED[4]) to detect the pectoral muscle boundary.

2 Methodology

2.1 Data

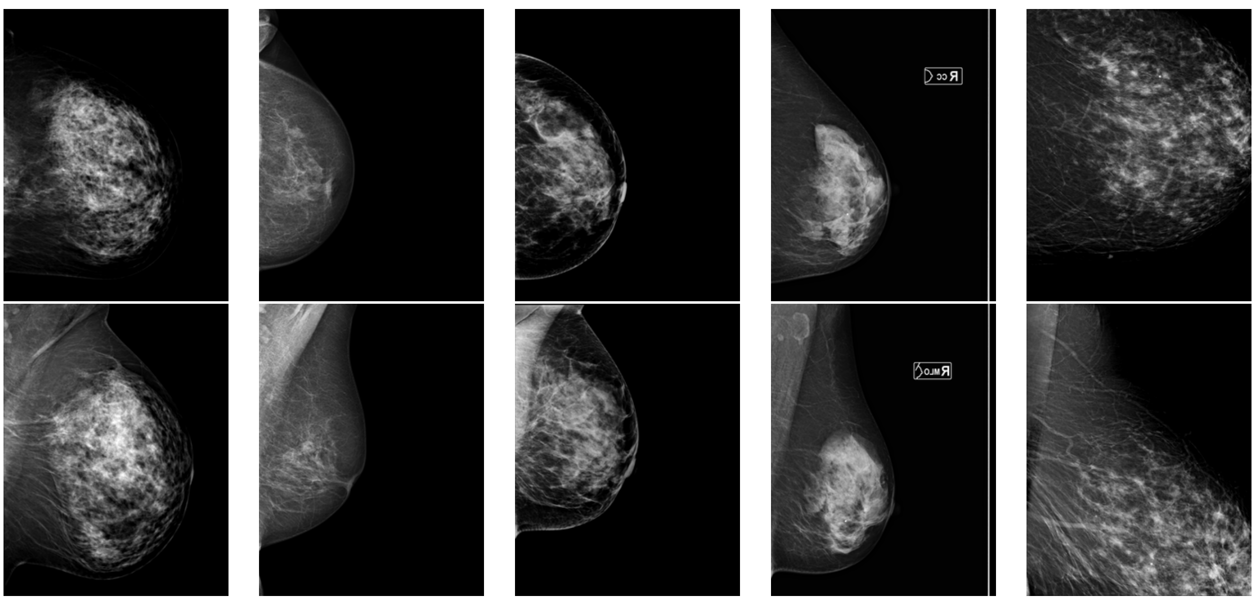
The aim of this work is to train a single network to do one-shot segmentation of background, breast tissue and pectoral muscle in FFDM and DBT images. Three datasets were used, which are detailed below. Each dataset has both craniocaudal (CC) and mediolateral-oblique (MLO) views.

2.1.1 *Multi-vendor FFDM*

This dataset consists of a large number of digital mammograms from five different vendors, distributed as listed in table 1. The cases for each vendor were split up into five folds and used for cross-validation. The DICOM images are preprocessed by downsampling each one to 100 micron pixel resolution and applying the appropriate window level before exporting to PNG. Artifacts like labels and non-homogeneous backgrounds were retained. Right oriented breasts were flipped to left orientation in order to homogenise the dataset.

Vendor	**Cases**	**Images (MLO)**
Fujifilm	995	3748 (1818)
Hologic	433	1628 (853)
GE	336	1579 (757)
Philips	400	1549 (761)
Siemens	553	2348 (1171)
Total	**2717**	**10852 (5360)**

Table 1: Overview of number of cases and images per vendor in the FFDM dataset.



(a) GE. (b) Siemens. (c) Hologic. (d) Fujifilm. (e) Philips.

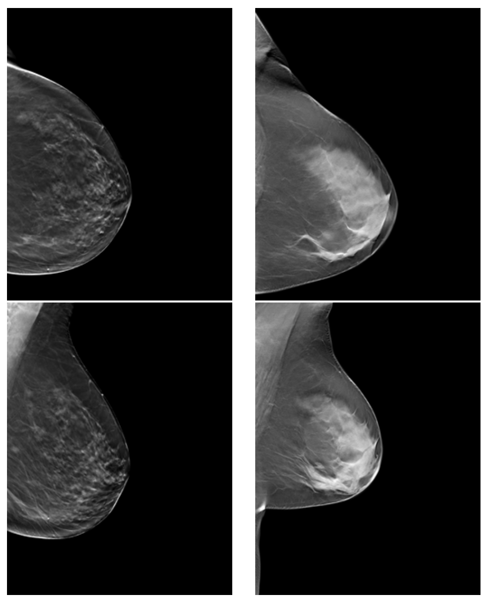
Fig 1: Examples of mammograms taken from different vendors. Top row depicts craniocaudal view of the breast, bottom row depicts mediolateral-oblique view of the same breast.

2.1.2 *Multi-vendor DBT*

This dataset consists of DBT images from two different vendors, distributed as listed in table 2. From each 3D volume, the central slice was selected to allow transfer of a 2D segmentation network trained on FFDM. Preprocessing was identical to that of the FFDM dataset. This dataset was only used for evaluation.

Vendor	Cases	Images (MLO)
Hologic	442	1725 (864)
Siemens	423	1822 (923)
Total	**865**	**3547 (1787)**

Table 2: Overview of number of cases and images per vendor in the DBT dataset.



(a) Hologic. (b) Siemens.

Fig 2: Examples of DBT center slice images from different vendors. Top row depicts craniocaudal view of the breast, bottom row depicts mediolateral-oblique view of the same breast.

2.1.3 *INbreast*

To compare results with existing literature, the INbreast dataset[5] is also included. The dataset includes 108 cases and 410 total images of which 206 are in the MLO view. This dataset was only used for evaluation.

2.2 *Network architectures*

U-Net has been repeatedly shown to be very successful in medical image segmentation tasks.[6] Initial experiments already showed very promising results for our task, but some common mistakes could be observed. For example, the network would predict pectoral muscle in unlikely places/shapes. To solve this, the network should be incentivised to pay special *attention* to salient features. In order to realise this idea, we took inspiration from Oktay et al.[7] and introduced attention gates to our U-Net architecture. These gates allow for filtering of higher level features using lower level features and learned attention coefficients.

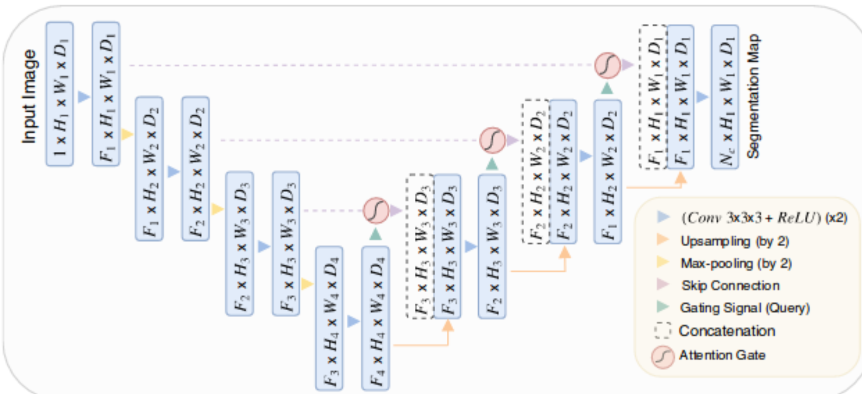


Fig 3: Attention U-Net architecture diagram.[7]

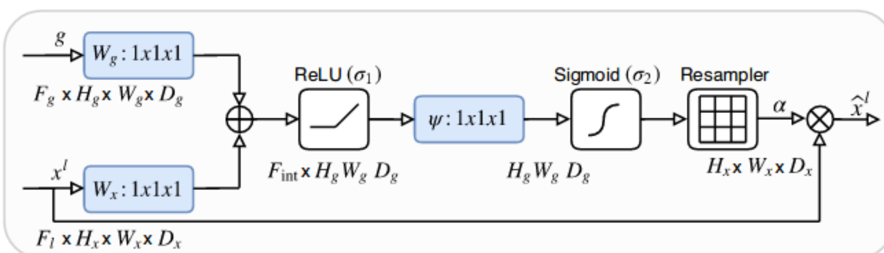


Fig 4: Attention gate used in Attention U-Net architecture from fig. 3.[7]

2.3 *Training*

All networks were trained using the Dice loss. An Adam optimiser with a learning rate of $5 \cdot 10^{-5}$ and L2 regularisation factor 10^{-2} was used. The ideal input image resolution was found to be 512×384 pixels, e.g. 400 micron pixel resolution, from our experiments trying 200–1600 micron resolutions.

3 Experiments & Results

We test the added benefit of introducing attention gates to the network. U-Nets with and without attention are trained on FFDM data only. The networks are evaluated on FFDM through cross-validation, and average scores of these models are reported on DBT and INbreast. The results of these experiments can be found in table 3. The scores reported for FFDM data are obtained using cross-validation

	FFDM	DBT	INbreast
Attention U-Net	**0.9848 (0.9453)**	**0.9809 (0.9242)**	0.9780 (0.7890)
U-Net	0.9836 (0.9453)	0.9763 (0.9084)	**0.9781 (0.8081)**

Table 3: Dice scores of both models on each dataset. The mean Dice score is listed first, and the pectoralis-specific Dice score is listed between parentheses.

Some good and bad examples of predicted segmentations can be found in fig. 5. The bottom-left image depicts the type of mistake that is usually visible with low Dice scores. In the top-right, the network's ability to deal with e.g. labels is demonstrated.

4 Discussion

In this work, we have shown that an attention gated U-Net can provide robust breast region segmentations in a multimodal multi-vendor setting. We observe that the hardest part of this task is predicting *plausible* segmentations. For instance some segmentations contains holes in the pectoral muscle, whereas for others the breast area contains islands of pectoral muscle.

We observe that the Dice score for the pectoral muscle was relatively low in the INbreast dataset compared to the other datasets. We believe this is due to the lack of contrast in the images. The images included in the INbreast dataset are much older than the in-house FFDM data, so they are expected to be quite different. We find that the Dice score jumps up to 0.95 when we train our method on the INbreast data. This may imply a lack of generalisation, which might be solved by e.g. contrast or gamma augmentation during training.

5 Future work

There are many approaches still to be explored. For instance, some methods exist that could teach a model to take "common knowledge" like plausible shapes into account. The problem of predicting little islands of tissue at atypical locations can be completely avoided by not trying to create dense segmentation maps, but instead estimating the pectoral muscle or skin-air boundaries as a line (as often done in conventional methods). This could be accomplished through edge detection networks. More intuitive penalisation in the loss function could also yield some performance gain, for example through a generalised Wasserstein Dice loss[8] where it is possible to specify inter-class relationships.

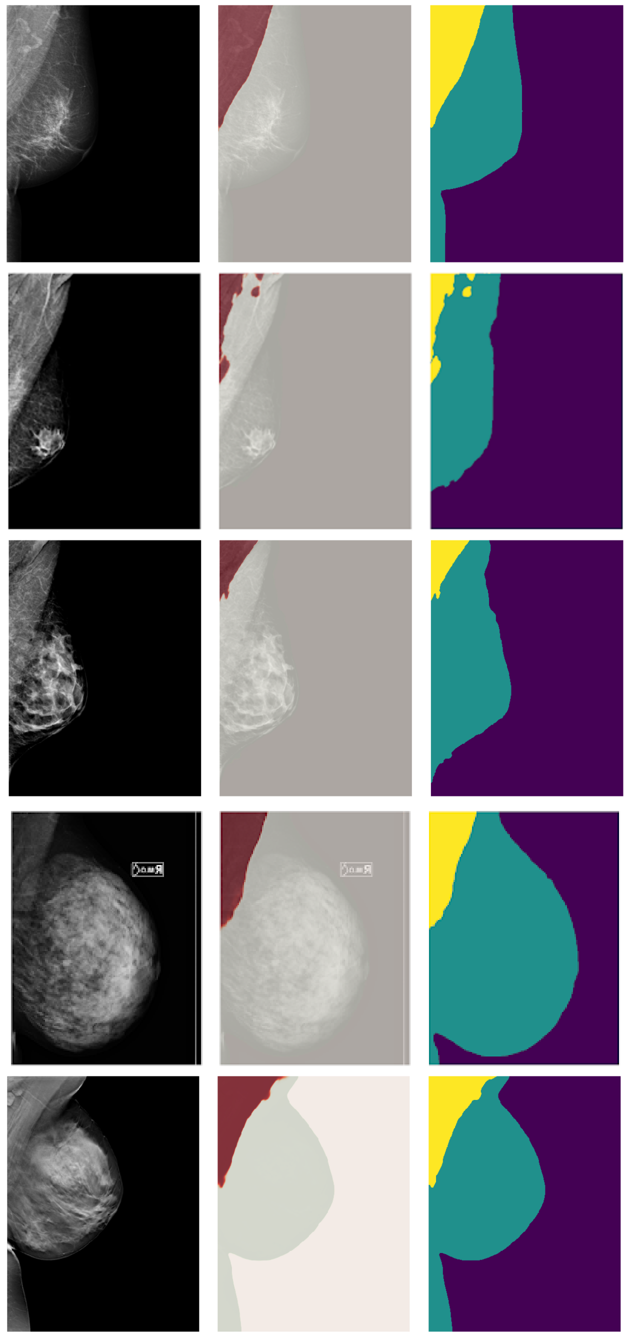


Fig 5: Some examples of predicted segmentations on FFDM (row 1-4) and DBT (row 5), showing the strengths and weaknesses of the proposed method. Left column depicts the original image, middle column depicts the class probability map, right column depicts the thresholded final mask. Some mistakes are still made due to skin folds or other artifacts.

References

1 M. Mustra, M. Grgic, and R. M. Rangayyan, “Review of recent advances in segmentation of the breast boundary and the pectoral muscle in mammograms,” *Medical & biological engineering & computing* **54**(7), 1003–1024 (2016).

2 A. Rodriguez-Ruiz, J. Teuwen, K. Chung, *et al.*, “Pectoral muscle segmentation in breast tomosynthesis with deep learning,” in *Medical Imaging 2018: Computer-Aided Diagnosis*, **10575**, 105752J, International Society for Optics and Photonics (2018).

3 A. Rampun, P. J. Morrow, B. W. Scotney, *et al.*, “Fully automated breast boundary and pectoral muscle segmentation in mammograms,” *Artificial intelligence in medicine* **79**, 28–41 (2017).

4 S. Xie and Z. Tu, “Holistically-nested edge detection,” in *Proceedings of the IEEE international conference on computer vision*, 1395–1403 (2015).

5 I. C. Moreira, I. Amaral, I. Domingues, *et al.*, “INbreast: toward a full-field digital mammographic database,” *Academic radiology* **19**(2), 236–248 (2012).

6 O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional networks for biomedical image segmentation,” in *International Conference on Medical image computing and computer-assisted intervention*, 234–241, Springer (2015).

7 O. Oktay, J. Schlemper, L. L. Folgoc, *et al.*, “Attention U-Net: Learning where to look for the pancreas,” in *Medical Imaging with Deep Learning*, (2018).

8 L. Fidon, W. Li, L. C. Garcia-Peraza-Herrera, *et al.*, “Generalised Wasserstein dice score for imbalanced multi-class segmentation using holistic convolutional networks,” in *International MICCAI Brainlesion Workshop*, 64–76, Springer (2017).