Article

Fully Automatic Thoracic Cavity Segmentation in Dynamic Contrast Enhanced Breast MRI Using Deep Convolutional Neural Networks

Marco Berchiolli [1,*], Susann Wolfram [2,3], Wamadeva Balachandran [4] and Tat-Hean Gan [1]

[1] Brunel Innovation Centre, Brunel University London, Uxbridge UB8 3PH, UK; tat-hean.gan@brunel.ac.uk
[2] Healthcare Innovation Centre, Teesside University, Middlesbrough TS1 3BX, UK
[3] School of Kinesiology, University of Michigan, Ann Arbor, MI 48109, USA
[4] Electronic and Computer Engineering Department, Brunel University London, Uxbridge UB8 3PH, UK; wamadeva.balachandran@brunel.ac.uk
[*] Correspondence: marco.berchiolli@brunel.ac.uk

Abstract: Dynamic Contrast Enhanced Magnetic Resonance Imaging (DCE-MRI) is regarded as one of the main diagnostic tools for breast cancer. Several methodologies have been developed to automatically localize suspected malignant breast lesions. Changes in tissue appearance in response to the injection of the contrast agent (CA) are indicative of the presence of malignant breast lesions. However, these changes are extremely similar to the ones of internal organs, such as the heart. Thus, the task of chest cavity segmentation is necessary for the development of lesion detection. In this work, a data-efficient approach is proposed, to automatically segment breast MRI data. Specifically, a study on several UNet-like architectures (Dynamic UNet) based on ResNet is presented. Experiments quantify the impact of several additions to baseline models of varying depth, such as self-attention and the presence of a bottlenecked connection. The proposed methodology is demonstrated to outperform the current state of the art both in terms of data efficiency and in terms of similarity index when compared to manually segmented data.

Keywords: DCE-MRI; breast segmentation; deep learning

1. Introduction

Breast cancer is the most frequent cancer among females, amounting to 24% of all cancer occurrences in 2018 [1], accounting for 684,996 deaths in 2020 worldwide [2]. The World Health Organization indicates screening programs aimed at early detection as one of the key factors in reducing mortality [3]. Magnetic Resonance Imaging (MRI) is an increasingly popular procedure for the screening of high-risk groups [4] and evaluating the response to neo-adjuvant chemotherapy [5]. MRI has several benefits over X-ray mammography; it does not utilize ionizing radiation, generates high-resolution images and contains dynamic information. Moreover, recent developments in MRI image processing [6] indicate that MRI is gaining popularity even with its major drawbacks (time-consuming, stressful and costly) and is actively being targeted by the research community.

Dynamic Contrast Enhanced MRI (DCE-MRI) outputs four-dimensional data (three spatial dimensions + one temporal dimension), consisting of images acquired before and after the intravenous injection of a contrast agent (CA). The change in tissue appearance in response to the CA is tissue-specific and, therefore, indicative of the presence of malignant breast lesions [7]. These changes in tissue appearance are extremely similar to the ones of the internal organs such as the heart, making automatic lesion detection in breast DCE-MRI sequences challenging. Manual delineation of the chest wall is an extremely time-consuming activity. Therefore, the automatic removal of the internal organs from the images by segmenting the chest cavity is instrumental to the development of an automatic

lesion detection methodology. Figure 1 shows the results of an automatic lesion detection algorithm for a properly segmented image and a poorly segmented image of a breast DCE-MRI sequence. The algorithm is based on the statistical properties of the whole sequence and indicates areas with a high likelihood of containing a lesion, with red indicating the most likely candidate area. By not excluding the chest cavity, the system evaluates the heart as an area of high likelihood. As the chest cavity is segmented, the actual lesion is correctly highlighted.
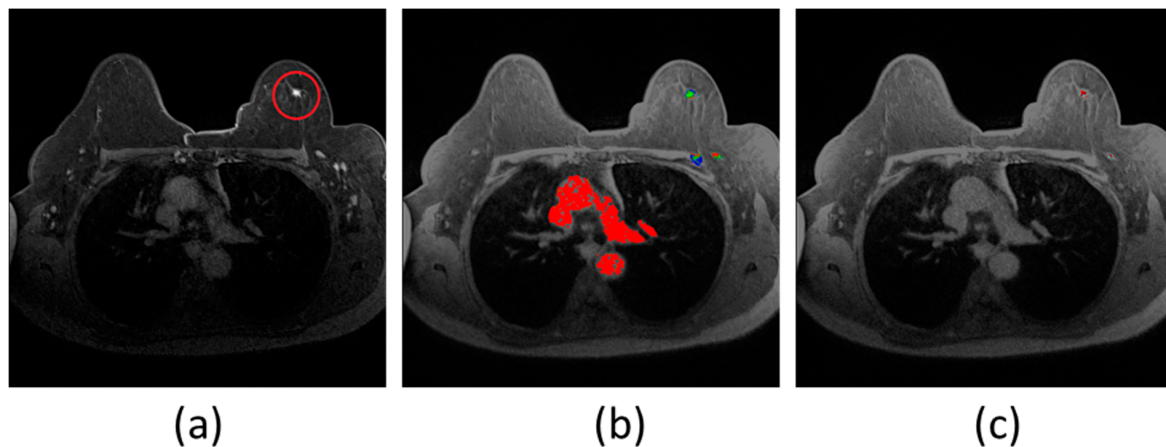
Figure 1. Results of an automatic lesion detection algorithm on a lesion. The original image with the lesion is visible in (**a**), with the lesion highlighted in the red circle. The coloring scheme of the other images (**b**,**c**) is based on statistical properties of the full breast volume. Red represents high likelihood of a lesion, while green and blue represent lower likelihood. In (**b**), the chest cavity is incorrectly segmented, and the lesion detection algorithm identifies the heart as an area of high likelihood to be a lesion. In (**c**), the chest cavity is correctly segmented, and the area of high likelihood to contain a lesion is correctly identified.

The current state of the art addresses the challenge by training a three-dimensional cluster of 2D UNets [8]. We argue that implementing more recent deep learning (DL) techniques can lead to better results and generalization performance.

In this study, a novel, DL-based methodology for the segmentation of the chest cavity from breast DCE-MRI scans is proposed. The solution aims to address the main challenges in chest cavity segmentation, especially sternum detection, by using a data-based approach. To showcase the potential of the solution, the target area to segment is selected to be the upper half of the chest cavity.

The paper is organized as follows. In Section 2, a selection of previous studies with relative contributions is presented; in Section 3, the proposal is presented in detail; in Section 4, the results of the experiments are presented and discussed; finally, in Section 6, conclusions and suggestions for further work are presented.

2. Related Work

Given the importance of the task and its complexity, there have been numerous approaches proposed in the published literature. In line with Marrone et al. [9], DL-based approaches are also included in the following overview. While the DL category has limited representation at present, it currently represents the state of the art. Moreover, it is expected that the popularity of this subfield will increase due to the vast representation of DL-based techniques in parallel fields, such as hand and brain segmentation [10]. For these reasons, the section dedicated to DL approaches is given higher importance.

2.1. Pixel-Based Approaches

Approaches in this category rely on classifying pixels or voxels individually, or with simple computations on the surrounding pixels [11,12]. Results are not always fully

automatic and tend to be suboptimal, especially the boundary between the sternum and the internal organ, which is often wrongly segmented. On the other hand, they require minimal computational costs. For example, in the approach of Vignati et al. [13], the images are processed using Otsu's thresholding and a sequence of dilations and erosions. Results show good breast parenchyma segmentation performance. However, the limitations become apparent as the examples provided show imprecise chest wall segmentation, as specified by the authors, and require the aid of fat-saturated images or an atlas-based segmentation. The study demonstrates that segmentation of the outer boundary of the breast is achievable by computationally efficient methodologies, while chest cavity segmentation would require more complex solutions. Given the importance of minimizing the presence of internal organs (see Figure 1), specifically the heart, chest cavity segmentation is fundamental in the development of lesion detection systems for DCE-MRI. Future studies should then focus on the detection of the chest wall.

2.2. Atlas-Based Approaches

The solutions in this category are generated by comparing the anatomical atlas generated by manually segmented data [12,14–17]. The approaches usually require a high number of instances in the atlas to guarantee generalizability to different anatomical features and acquisition protocols. The size of the atlas, however, is directly linked to an increased computational cost. To counteract such limitations, the solutions are often restricted to a specific anatomical part. An example of such an application is provided by Fooladivanda et al. [18], in which the authors use an atlas-based approach to segment the pectoral muscle, relying on a simpler pixel-based approach to segment the chest wall edge.

2.3. Geometrical-Based Approaches

The proposed solutions within the category revolve around constraining the segmentation results to predetermined anatomical and physiological characteristics. Their most common usage is as a refinement to pixel-based approaches [19,20]. The main criticisms of the techniques are the extreme computational costs and extremely poor generalization performance. Notable examples come from Wu et al. [20], in which the authors propose a methodology to extract the chest wall line from sagittal breast MRI.

Results are based on the refinement of edge detection via enforcing geometrical constraints and are extremely effective. The fundamental assumptions, however, highlight the limitations in generalizability, as the methodology does not account for the edge detection algorithm failing. This limits the potential application in a contrast-enhanced MRI environment, as post-enhancement images usually feature high-intensity chest wall and heart regions, as shown in Figure 2.
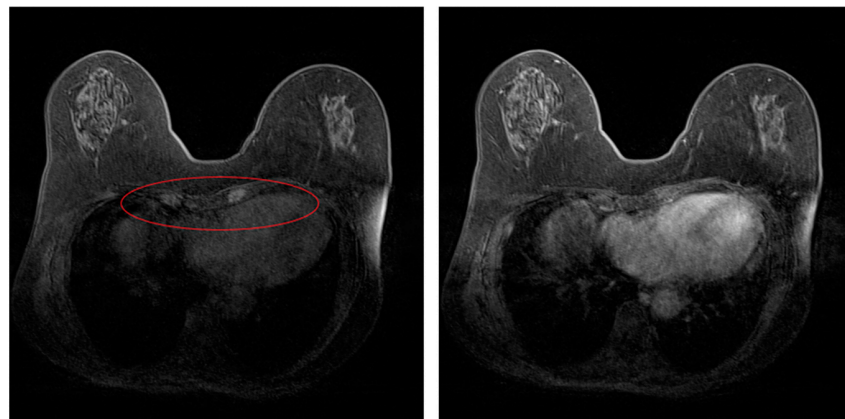
Figure 2. Pre-enhancement and post-enhancement (6 min after CA injection) images of a central slice of the breast. The area of the sternum highlighted in the picture on the left is not clearly defined through an edge detection algorithm in the picture on the right.

2.4. Deep Learning-Based Approaches

Deep learning-based solutions allow for the creation of fully automated segmentation methodologies by exploiting previous examples in the form of manually segmented data. The limitations of the approaches consist in the training process, for which large quantities of data and computational resources are needed. The techniques have been applied with great success to an increasing amount of research problems in recent years [21].

The pioneering solution is attributable to Dalmis et al. [22], in which the authors propose automated breast and fibroglandular tissue (FGT) segmentation using a UNet approach [23]. They obtained considerable improvements in accuracy over the state of the art at that time in terms of Dice Similarity Coefficient (DSC), achieving a DSC of 0.944 against 0.863 from the previous reported state of the art [12,15]. The current state of the art is attributable to Piantadosi et al. [8], who implemented a multi-planar UNet approach, obtaining some improvements over Dalmis et al. [23]. The methodology uses three discrete UNets on the transverse, sagittal and coronal planes instead of a three-dimensional (3D) UNet, to increase the computational efficiency for the solution. The 3D aspect of the network, however, effectively triples the computational cost. Moreover, the current state of the art [8] was trained on a dataset composed of fully labeled data from 117 patients, well above the data available to many developers. As the labeling process is the most time-consuming aspect of the development, solutions should aim at maximizing performance with a minimal amount of data. To this end, novel techniques need to be employed, both architecturally and from a data pre-processing point of view.

Given the limitations of the approaches presented and their advantages, the convolutional neural network (CNN) methodologies are the most applicable for problems in computer-aided diagnostics (CAD). However, all DL approaches available in the literature fail to address the most limiting aspect of the issue, which is data availability. The proposed solution aims to improve the current state of the art by focusing on data efficiency, by employing architectures that are less prone to overfitting and behave better in transfer learning scenarios. Moreover, recent advancements in DL research, such as self-attention [24], have the potential to further improve upon the state of the art. In addition, computational efforts need to be compliant with the hardware that is potentially available in a hospital; hence, the algorithms should process patient data in a two-dimensional fashion, as 3D data can be generated if needed.

The current paper reports on improvements to the current state of the art by employing several techniques that have been successful in parallel fields. To this end, a total of 18 configurations were evaluated, corresponding to the combination of three different architectures and three different techniques to be applied.

3. Methodology

3.1. Introduction

The proposed automatic chest cavity segmentation algorithm is composed of two main parts. In the first part, the volume is sliced in its transverse plane images, and it is inputted in a Dynamic UNet (fast.ai, n.d.), inspired by [25].

Figure 3 shows the schematic of the solution, with an image being inputted in the DL model to generate a mask. The mask is then applied to the input image.

3.2. Dataset and Labeling Strategy

A dataset consisting of breast DCE-MRI data from 44 patients was used for the training and testing of the proposed segmentation model. Data were acquired on a 1.5 T scanner (MAGNETOM Avanto, Siemens Healthcare GmbH, Erlangen, Germany) with the patient positioned lying face down. The TR/TE/flip angle was 4.33 s/1.32 s/10° for each scan, with a slice thickness of 1.1 mm, with no gaps between slices. The resolution of each slice was 448×448 pixels. Each breast DCE-MRI protocol consisted of one pre-contrast T1-weighted sequence and seven post-contrast T1-weighted sequences collected at intervals of 1:01 min between sequences.
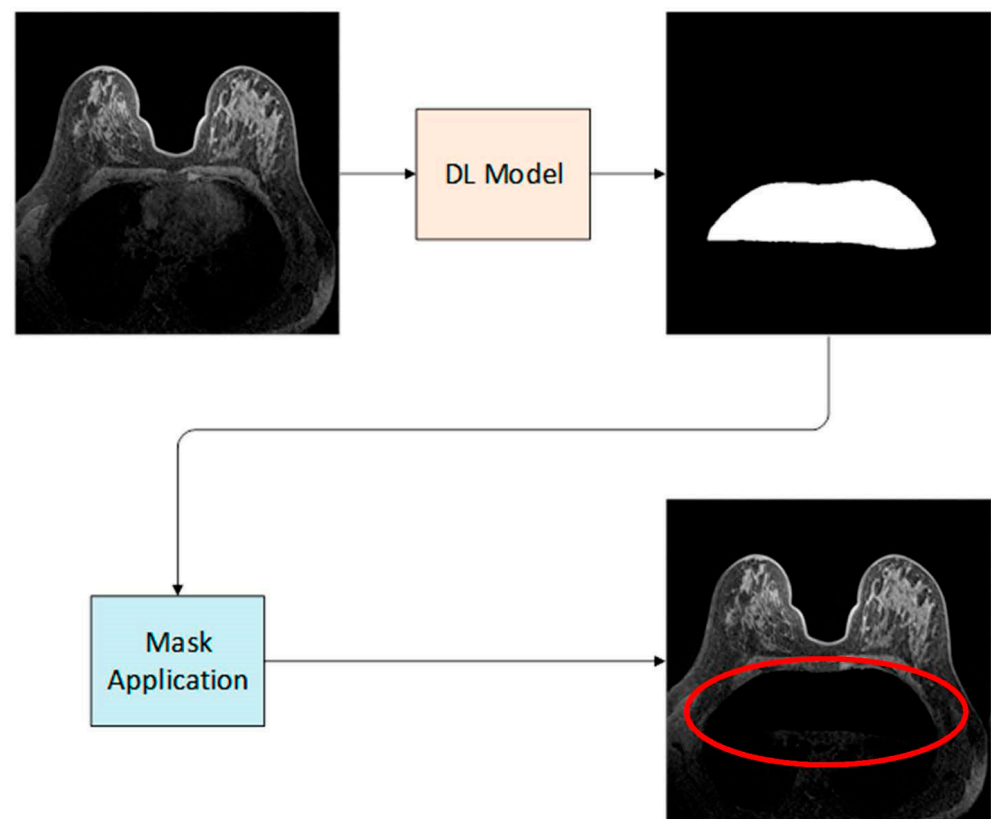
Figure 3. Schematic of the proposed solution. The input image is inputted in a deep learning model, which outputs a generated mask. The segmentation mask is then applied to the input image. Notably, the upper part of the internal organs is removed, as highlighted by the red circle.

A subset of the slices of each sequence was manually segmented, and the segmentation was checked for accuracy by a breast radiologist. Before manual segmentation, the slices were cropped to size to remove the posterior half of the thorax. In this developmental phase of the segmentation, the bottom row of 131 pixels was removed as this gave the best cropping result for our dataset. For manual segmentation, eleven evenly spaced slices throughout the central 70 slices of the scan volume of the pre-contrast sequence were selected to address anatomical variations along the upper body. To address changes in tissue brightness because of CA injection, the six central slices of the pre-contrast sequence were also manually segmented. The manual segmentation of these slices was repeated for the corresponding slices in all seven post-contrast sequences. A total of 2552 slices were manually segmented. Segmented slices from 37 patients (n = 2146 slices) were used for the training of the model. The model was tested with segmented slices from the remaining seven patients (n = 406 slices).

3.3. Deep Learning Model

The segmentation model is based on a Dynamic UNet with a pre-trained ResNet encoder. A transfer learning approach was utilized with a pre-trained ResNet encoder to address the challenge of insufficient training data. Transfer learning [26] is a widely used technique in computer vision tasks outside of medical image segmentation [27,28]. In order to leverage transfer learning, the technique employed in the proposed solution is the Dynamic UNet (fast.ai, n.d.), based on the original Unet proposed by Ronneberger et al. [23]. A UNet architecture was chosen as it is the best methodology in terms of overall performance in medical imaging applications [21,29]. The flexibility derived from using custom encoders provided by Dynamic UNet allows for a much greater degree of exploitation of transfer learning.

The original UNet architecture can be divided into an encoding part, or "downsampling", and a decoding part, or "upsampling". The encoding side performs a similar task to a conventional CNN, with regular downsampling steps performed through the maxpool operation. At the same time, the decoding path follows a symmetrical structure, with the upsampling steps performed through a fractionally strided convolution layer. The symmetry allows for the activations of the downsampling layers to be concatenated to the activations of the upsampling layers. This offers better retention of spatial information throughout the upsampling path.

The Dynamic UNet architecture was originally presented by Iglovikov and Shvets [25]. It follows the same basic architecture as the original UNet but adds a pre-trained model as the encoder. The approach not only achieves considerable improvements over the traditional UNet, but it also allows one to experiment with a variety of pre-trained encoders. In this study, the ResNet encoders were used, as it has been demonstrated that they reduce the reliance on regularization techniques [30].

ResNets were first introduced in 2015 [31] to counteract the vanishing/exploding gradient problem in deeper networks, which had been a mainstay challenge in the deep learning research field since the inception of ConvNets [32]. The fundamental component of a ResNet is the residual block, which incorporates a "skip connection". The input to the block is passed through an identity mapping and is summed to the activation of a series of convolutional layers. ResNets can reach theoretically infinite depths, due to the self-regularization method provided by the identity mapping [31]. The flexibility of ResNets allows training for longer (more epochs), thus reducing the likelihood of overfitting [31].

3.4. Model Configuration

In addition to the novel application of the Dynamic UNet in this field, several architectural configurations were added to the models. These model configurations further improve upon the current state of the art. The configurations introduced to the model were (Figure 4).
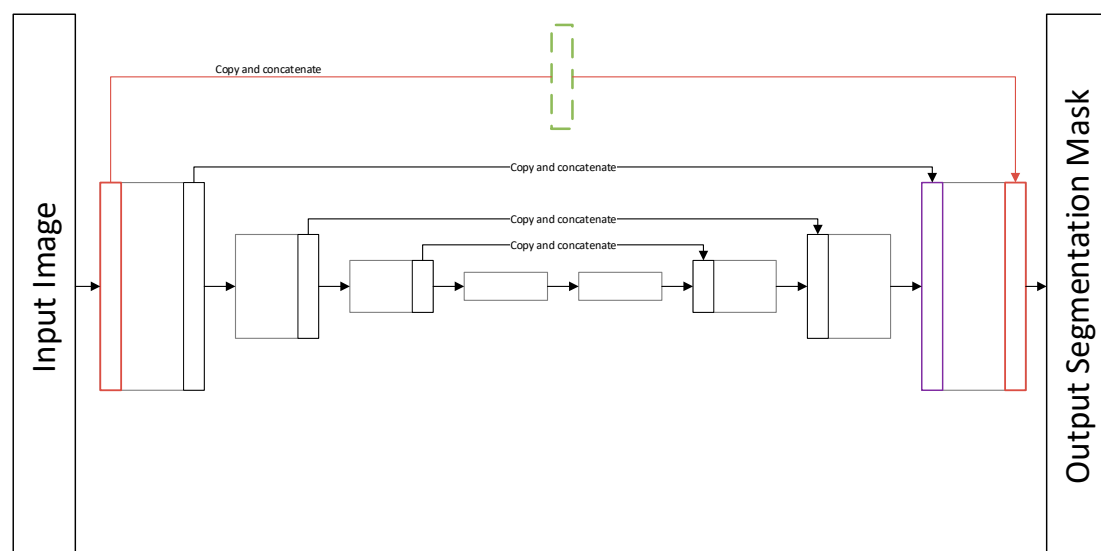
Figure 4. Experimental layout. Purple represents the location of the self-attention layer (here represented in a simplified view of a ResNet18). Red represents the skip connection from the first to last layer. Green is the bottlenecked connection.

- ResNet encoders;
- A self-attention layer as part of the upsampling path of the model;
- A blurring algorithm to avoid checkerboard artefacts;
- A bottleneck connection from input to output.

3.4.1. ResNet Encoders

The chosen encoder architectures were ResNet18, ResNet34 and ResNet50 [31]. It was important to convey the effect of architectural scaling on performance, and ResNets have been demonstrated to be an excellent choice for medical imaging work [33,34]. While the number of training parameters is considerably higher than in most other solutions, the increased transferability of ResNets allows for improved results with lower amounts of data, thus validating the effort that was put into increasing data efficiency. The choice of not experimenting with larger architectures, such as ResNet101, was made due to insufficient data for such architectures. This assumption was confirmed correct by the results.

3.4.2. Self-Attention Layer

Attention is a mechanism that was introduced by Bahdanau et al. [35] with the aim of improving neural machine translation tasks. Attention layers have since been an integral part of transformer-based models [24]. Attention-based models have been shown to excel in all contexts in which capturing global dependencies is necessary, including hybrid text–image tasks [36,37] and computer vision tasks [38]. Self-attention consists of a block of layers that outputs the attention feature maps of an input sequence with each element of the same sequence. Self-attention has been featured in promising research within CAD [39], with the usage of self-attention as a region-of-interest detection tool, thus allowing one to obtain a cropped local image in which to perform a classification. The experiments were run with the implementation described by Zhang et al. [40], who introduced a ResNet-like approach to self-attention by including a skip connection within the final output of a layer y_i:

$$y_i = \gamma o_i + x_i$$

where γ is a learnable parameter that scales the output of a self-attention layer and x_i is the input of the layer. The approach results in greater spatial awareness by the model, a trait that is highly desirable in medical imaging, as images often present themselves with a pseudo-symmetrical and repetitive structure.

3.4.3. Blurring

To counteract the natural occurrence of checkerboard artefacts in CNNs [41], a blurring mechanism was introduced. An average pooling layer with 2×2 dimensions and the unit stride was added after each activation in the upsampling path of the Dynamic UNet.

3.4.4. Bottlenecked Connection

The original UNet architecture did not feature any direct connection between the input and the final layer, opting instead for a concatenation of the activation of the third layer and the activation of the last deconvolution layer [23]. Adding an even less processed pass-through could lead to better spatial awareness and improved performance. In this work, a bottlenecked connection is included, aiming at forcing the model to synthetize the input information by using a bottleneck within the residual block, which halves the number of features in the convolutional path of the block.

3.5. Training of the Models

3.5.1. Data Augmentation

In addition to transfer learning, we further address the issue of insufficient training data by employing a more aggressive data augmentation strategy compared to the published literature. Data augmentation is an ensemble of techniques that are employed to artificially increase the amount of available data with the aim of reducing overfitting and improving generalization [42]. Specifically, in computer vision, images are slightly altered through affine or lighting transformations. These can include rotations, cropping, contrast and color correction. By heavily employing data augmentation, the chances of overfitting are drastically lowered, allowing for improved performance with data being equal (by

employing a larger architecture), or by achieving similar results with lower amounts of data, as presented by Wong et al. [43].

We employed several augmentation strategies (summarized in Table 1), and augmented images were used for all model configurations. All transformations were performed with a reflection padding mode, mirroring the pixel values along the image border to fill the shape. Examples of augmented data can be seen in Figure 5. The combined probability of a transformation to a specific feature is

$$p = 1 - \prod (1 - p_i)$$

where p_i is the probability of the transformations that affect the specific feature to occur. The dataset was then composed of ~99.95% data-augmented images. The perfectly horizontal edge of the manually segmented mask was impacted by the perspective warp and the rotation transformation. This led to an overwhelming imbalance in the dataset, with ~93.75% of the images featuring an inclined mask as ground truth.

Table 1. Data augmentation overview. The probability column refers to the likelihood of any transformation occurring.

Transformation	Parameters	Probability
Horizontal flip	N/A	0.5
Rotation	±10°	0.75
Cropping	1.1 magnification	0.75
Contrast adjustment	±20%	0.75
Brightness adjustment	±10%	0.75
Perspective warp	±20% position of the observation plane	0.75
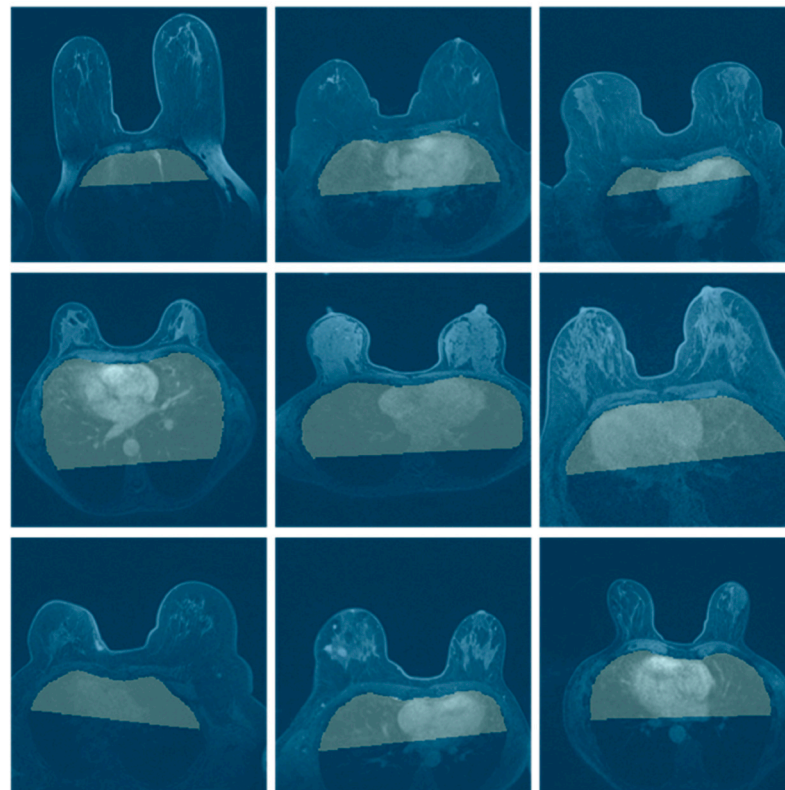
Figure 5. Example of data augmentation on a small batch of data. The reflection padding can be seen in the top left corner. All images have their manually labeled ground truth featuring a straight line at the bottom of the mask.

3.5.2. Hyperparameters

To identify the best segmentation model for the segmentation task, a total of 18 different segmentation models were trained, each with a different combination of the configurations described above. The same training datasets (2146 images from 37 patients) were used for the training of all model configurations. Every model was trained with three distinct random seeds to ensure minimal stochastic noise in the results.

The optimizer for the training was Adam [44], with β_1 of 0.9 and β_2 of 0.99. The optimal learning rate for each architecture was found as shown by Smith [45], resulting in an optimal theorical learning rate of 1×10^{-4} for all configurations. The training phase featured learning rate annealing, as described by the 1-cycle policy [46]. All trainings were performed with weight decay of 0.01.

The models were trained using an NVIDIA P40 with 24 GB of VRAM. Batch sizes were chosen empirically, with ResNet18 models having 64 (31,208,178 parameters), ResNet34 having 32 (41,316,338 parameters) and ResNet50 having 8 (341,254,226 parameters).

3.5.3. Experiments

The performance of each of the 18 different models was tested with images from seven patients not used for training. The computational demand of each model was evaluated by calculating inference times for segmentation. Inference times were calculated for each individual image and for all images of the DCE-MRI sequence of each patient (n = 1260).

The segmentation result of each model was compared to 406 images (58 images per patient) that were manually segmented. Comparisons of agreement between the model segmentation and the manual segmentation were made using the DSC and the Jaccard Similarity Coefficient (JSC):

$$DSC = \frac{2|X \cap Y|}{|X| + |Y|}$$
$$JSC = \frac{|X \cap Y|}{|X \cup Y|}$$

where X is the number of pixels inside the thoracic cavity that are part of the manual segmentation and Y is the number of pixels inside the thoracic cavity that are part of the model segmentation. The DSC represents the mean overlap and the JSC represents the union overlap of pixels that are common to both the manual and automatic segmentation. For each model, DSC and JSC were determined for each image and the mean $\pm$ standard deviation (SD) of each coefficient was calculated for all 406 images.

4. Results

The DSC and JSC for each model are summarized in Table 2. Inference times are summarized in Table 3. The best agreement between model segmentation and manual segmentation was found for the ResNet34 model configuration with self-attention, the bottlenecked connection and the blurring mechanism. It achieved a DSC of 0.9359, JSC of 0.8874 and inference times of 33.56 ms/image and 42.3 s for all 1260 images of one DCE sequence. The worst agreement between model segmentation and manual segmentation was found for the ResNet50 with self-attention model configuration, with a DSC of 0.9210, JSC of 0.8670 and inference times of 194.8 ms/image and 245.2 s for all 1260 images of one DCE sequence.

Upon visual inspection of the model segmentation results, it was found in all models that the lower edge of the segmented area was not straight, which influenced the results of the similarity coefficients. This phenomenon was an artefact caused by the data augmentation strategy, which allowed for reduced overfitting and better training outcomes. The training data were rarely not augmented before being inputted into the models, resulting in the horizontal line at the bottom of the manually segmented mask being interpreted as always inclined. A simple adjustment step was then taken to ensure the real-world usability metrics of the methodology. The refinement algorithm consists of a simple loop over the columns of the output of the model, forcing any pixel between the highest true value in a column and the 131st from the bottom to be true. In other words, the generated

masks are refined to only allow for pixels above the height of 131 from the bottom of the image to be true. The updated schematic of the solution can be seen below in Figure 6.

Table 2. Similarity coefficients for all models. The similarity coefficients are represented as mean $\pm$ standard deviation. The best performing model was the ResNet34 model with a self-attention layer, bottleneck connection and blurring mechanism, which is highlighted in bold. Abbreviations: BC—bottleneck connection, BL—blurring mechanism, DSC—DSC Similarity Coefficient, JSC—Jaccard Similarity Coefficient, SA—self-attention layer, SD—standard deviation.

Model Configuration	Mean DSC $\pm$ SD	Mean JSC $\pm$ SD
Resnet18	0.9253 ± 0.1034	0.8760 ± 0.0758
ResNet18 + SA	0.9296 ± 0.1028	0.8757 ± 0.0765
ResNet18 + BL	0.9273 ± 0.1033	0.8795 ± 0.0764
ResNet18 + BC	0.9283 ± 0.1045	0.8742 ± 0.0763
ResNet18 + SA + BL	0.9348 ± 0.1045	0.8846 ± 0.0760
ResNet18 + SA +BL + BC	0.9293 ± 0.1036	0.8755 ± 0.0765
Resnet34	0.9244 ± 0.1017	0.8721 ± 0.0756
ResNet34 + SA	0.9230 ± 0.1012	0.8652 ± 0.0752
ResNet34 + BL	0.9227 ± 0.1015	0.8714 ± 0.0749
ResNet34 + BC	0.9292 ± 0.1009	0.8754 ± 0.0755
ResNet34 + SA + BL	0.9337 ± 0.1008	0.8780 ± 0.0750
ResNet34 + SA + BL + BC	$\mathbf{0.9359 \pm 0.1004}$	$\mathbf{0.8874 \pm 0.0748}$
Resnet50	0.9240 ± 0.1055	0.8717 ± 0.0781
ResNet50 + SA	0.9210 ± 0.1069	0.8670 ± 0.0790
ResNet50 + BL	0.9233 ± 0.1063	0.8708 ± 0.0777
ResNet50 + BC	0.9257 ± 0.1059	0.8730 ± 0.0775
ResNet50 + SA + BL	0.9278 ± 0.1061	0.8727 ± 0.0766
ResNet50 + SA +BL + BC	0.9289 ± 0.1053	0.8740 ± 0.0770

Table 3. Inference times for all models. Times are represented in ms/image for single image processing, and in seconds for batch processing. The best performing model was the ResNet18 with no additions, which is highlighted in bold. Abbreviations: BC—bottleneck connection, BL—blurring mechanism, DSC—DSC Similarity Coefficient, JSC—Jaccard Similarity Coefficient, SA—self-attention layer.

Model Configuration	Inference Time (ms/Image)	Batch Inference Time (1260 Images) (s)
Resnet18	**26.83**	**30.0**
ResNet18 + SA	33.56	42.3
ResNet18 + BL	31.88	40.2
ResNet18 + BC	27.68	34.9
ResNet18 + SA + BL	33.56	42.3
ResNet18 + SA +BL + BC	33.56	42.3
Resnet34	31.88	40.2
ResNet34 + SA	33.56	42.3
ResNet34 + BL	33.56	42.3
ResNet34 + BC	30.2	38.1
ResNet34 + SA + BL	33.56	42.3
ResNet34 + SA + BL + BC	33.56	42.3
Resnet50	179.5	226.2
ResNet50 + SA	194.6	245.2
ResNet50 + BL	196.3	247.3
ResNet50 + BC	194.6	245.2
ResNet50 + SA + BL	196.3	247.3
ResNet50 + SA +BL + BC	198	249.5

The sole purpose of the refinement algorithm is to remove the bias inserted into the training process by the aggressive data augmentation strategy. Such an algorithm

would not be used in a workflow for lesion detection, or in clinical practice, as it aims at improving performance in areas that result in no additional value in CAD or diagnosis. The performance metrics of the refined masks should then be interpreted as directly correlated with the actual capability of models to segment the chest cavity.
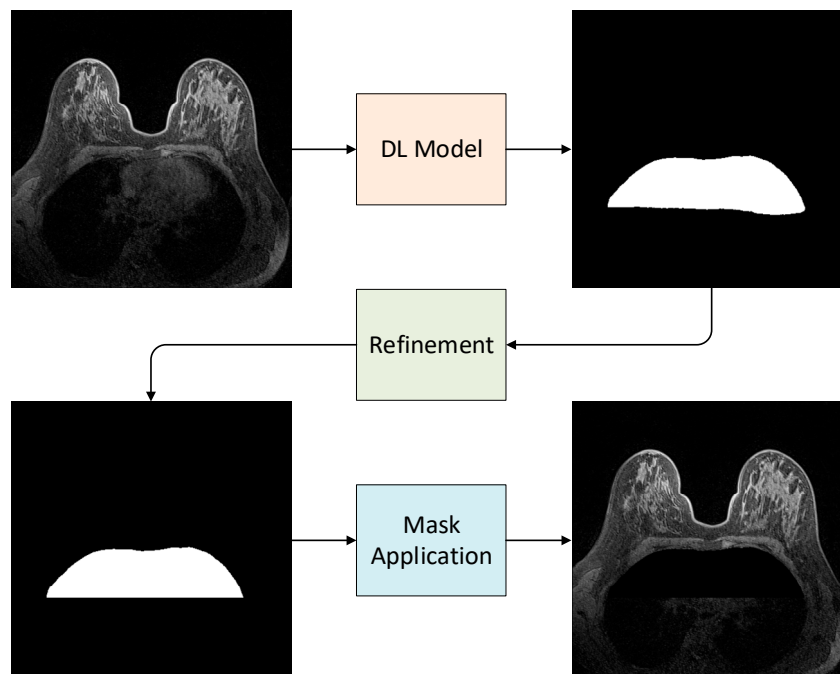
Figure 6. Schematic of the solution. The input image is inputted in a deep learning model, which outputs a generated mask. A refinement algorithm then removes the bottom part of the mask. The segmentation mask is then applied to the input image.

The application of the refinement algorithm led to considerable improvements in the similarity coefficients, but with a slight increase in inference times (Tables 4 and 5). The best agreement between model segmentation and manual segmentation was again found for the ResNet34 model with self-attention, bottlenecked connection and blurring mechanism model configuration. It achieved a DSC of 0.9612 and JSC of 0.9789, with inference times of 98.99 ms/image. The worst agreement between model segmentation and manual segmentation was found for the ResNet50 with self-attention model configuration, with a DSC of 0.9465 and JSC of 0.9708, with inference times of 261.7 ms/image.

Table 4. Similarity coefficients for all models after refinement algorithm. The similarity coefficients are represented as mean $\pm$ standard deviation. The best performing model was the ResNet34 model with a self-attention layer, bottleneck connection and blurring mechanism, which is highlighted in bold. Abbreviations: BC—bottleneck connection, BL—blurring mechanism, DSC—DSC Similarity Coefficient, JSC—Jaccard Similarity Coefficient, SA—self-attention layer.

Model Configuration	DSC $\pm$ SD (n = 406)	JSC $\pm$ SD (n = 406)
Resnet18	0.9750 $\pm$ 0.0451	0.9552 $\pm$ 0.0667
ResNet18 + SA	0.9731 $\pm$ 0.0449	0.9552 $\pm$ 0.0669
ResNet18 + BL	0.9739 $\pm$ 0.0447	0.9536 $\pm$ 0.0701
ResNet18 + BC	0.9717 $\pm$ 0.0450	0.9497 $\pm$ 0.0670
ResNet18 + SA + BL	0.9751 $\pm$ 0.0448	0.9556 $\pm$ 0.0665
ResNet18 + SA +BL + BC	0.9737 $\pm$ 0.0447	0.9533 $\pm$ 0.0674
Resnet34	0.9744 $\pm$ 0.0420	0.9541 $\pm$ 0.0682
ResNet34 + SA	0.9734 $\pm$ 0.0419	0.9527 $\pm$ 0.0628
ResNet34 + BL	0.9698 $\pm$ 0.0417	0.9512 $\pm$ 0.0642

Table 4. *Cont.*

Model Configuration	DSC ± SD (n = 406)	JSC ± SD (n = 406)
ResNet34 + BC	0.9766 ± 0.0422	0.9577 ± 0.0677
ResNet34 + SA + BL	0.9775 ± 0.0425	0.9584 ± 0.0633
ResNet34 + SA + BL + BC	**0.9789 ± 0.0411**	**0.9612 ± 0.0621**
Resnet50	0.9718 ± 0.0493	0.9541 ± 0.0627
ResNet50 + SA	0.9665 ± 0.0502	0.9538 ± 0.0721
ResNet50 + BL	0.9708 ± 0.0488	0.9465 ± 0.0704
ResNet50 + BC	0.9732 ± 0.0499	0.9526 ± 0.0706
ResNet50 + SA + BL	0.9712 ± 0.0501	0.9532 ± 0.0710
ResNet50 + SA +BL + BC	0.9766 ± 0.0497	0.9577 ± 0.0708

Table 5. Inference times for all models with subsequent refinement algorithm. Times are represented in ms/image for single image processing, and in seconds for batch processing. The best performing model was the ResNet18 with no additions, which is highlighted in bold. Abbreviations: BC—bottleneck connection, BL—blurring mechanism, DSC—DSC Similarity Coefficient, JSC—Jaccard Similarity Coefficient, SA—self-attention layer.

| Model Configuration	Inference Time (ms/Image)	Batch Inference Time (1260 Images) (s)
Resnet18	**92.28**	**116.3**
ResNet18 + SA	93.96	118.4
ResNet18 + BL	92.28	116.3
ResNet18 + BC	92.28	116.3
ResNet18 + SA + BL	95.64	120.5
ResNet18 + SA +BL + BC	95.64	120.5
Resnet34	90.6	114.2
ResNet34 + SA	95.64	120.5
ResNet34 + BL	95.64	120.5
ResNet34 + BC	92.28	116.3
ResNet34 + SA + BL	97.32	122.6
ResNet34 + SA + BL + BC	98.99	124.7
Resnet50	258.4	325.6
ResNet50 + SA	263.4	331.9
ResNet50 + BL	261.7	329.7
ResNet50 + BC	260.1	327.7
ResNet50 + SA + BL	261.7	329.7
ResNet50 + SA +BL + BC	266.8	336.2

5. Discussion

The main aim of this work was to develop a fully automatic methodology for the segmentation of the chest cavity to use for the development of automatic lesion detection systems. All 18 models showed excellent agreement with the manual segmentation, with DSCs of over 0.92 and JSCs of over 0.86. After the introduction of a refinement algorithm to compensate for artefacts due to the aggressive data augmentation strategy, the DSCs and JSCs improved to over 0.95 and 0.96, respectively. The best performing model was found to be the ResNet34 model with the self-attention layer, blurring mechanism and bottleneck connection, resulting in very high agreement with manual segmentation, with a DSC of 0.9612 and a JSC of 0.9789.

Overall, the techniques proposed show a significant improvement over the current state of the art [8], which features a DSC of 0.9660 on full breast segmentation. While the tasks are not directly comparable, body–air segmentation has been shown to achieve more than optimal results in the past [20], hence highlighting the need for novelty solely in chest cavity segmentation. Moreover, the DSC is expected to be lower with smaller regions to segment. Hence, it is reasonable to assume that the cost to the DSC and JSC of generating incorrect chest wall segmentation is lower, as shown in Piantadosi et al. [8].

The best configuration also improves inference timings, matching the previous state of the art after post-processing on comparable hardware, but dramatically improving on it without. As previously mentioned, the post-processing algorithm is solely used to highlight the validity of the proposed solution and is not meant to be included in any CAD workflow.

The agreement between the model segmentation and the manual segmentation was very good to excellent for all model configurations. The addition of each proposed improvement to the models provided marginal improvements for all architectures. However, the models with all configurations added consistently resulted in higher performance. This improvement was particularly striking for the ResNet 34 encoder, which showed an increase in the DSC from 0.9541 (ResNet34 only) to 0.9612 (ResNet34 with self-attention layer, blurring and bottleneck). However, adding to the complexity of the model comes with an increased computational cost, especially when adding the self-attention layer. For example, the inference time for the entire DCE image sequence is 114.2 s for ResNet34 only, but this increases by 6.3 s when adding the self-attention layer, while adding the bottleneck connection causes an increase of only 2.1 s (Table 5). The best performing model before post-processing is the ResNet34 featuring all proposed additions. The considerable improvement in segmentation accuracy over a standard ResNet encoder may justify the additional computational cost. ResNet50 configurations performed worse than the smaller ResNet34 architectures. This is likely due to underfitting, and having access to more data would be extremely beneficial and most likely would lead to better performance than ResNet34 on larger training datasets. The difference between the worst performing architecture (ResNet18 with bottleneck connection) and the best performing architecture (ResNet34 with all proposed additions) can be seen in Figure 7 below.

Input	ResNet18 + BC	ResNet34 + SA + BL + BC

(a)	(b)	(c)
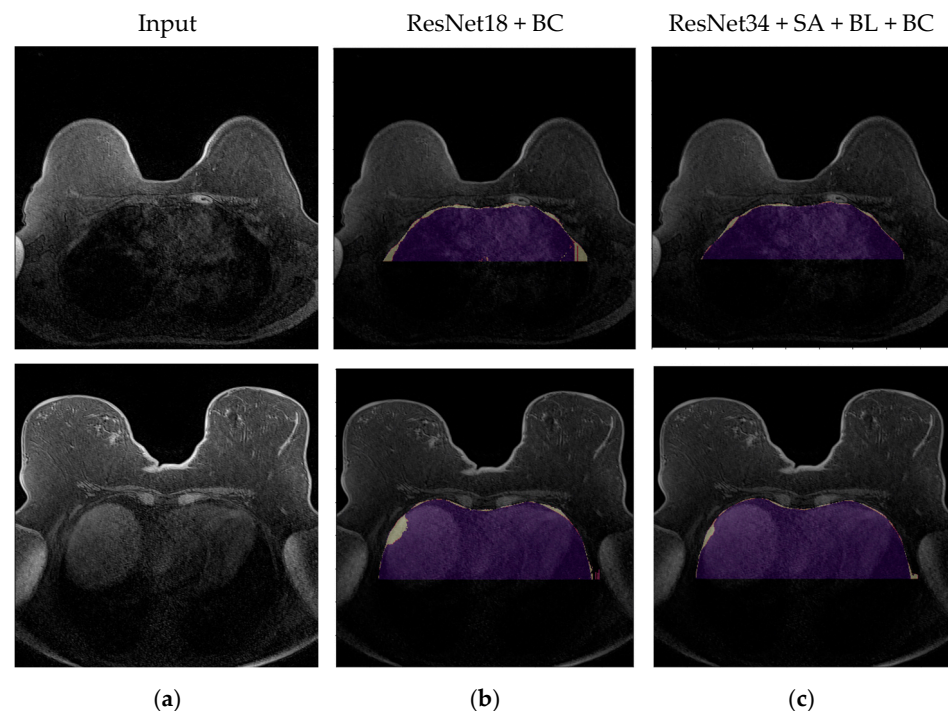
Figure 7. Comparison of results. The input images to the algorithms (**a**); the results from the worst performing algorithm (ResNet18 with bottleneck connection) (**b**); the results from the best performing architecture (**c**). The generated masks are purple, while the yellow coloring represents the areas of disagreement with the manual labeling. Abbreviations: BC—bottleneck connection, BL—blurring mechanism, SA—self-attention layer.

In cases of limited computational resources, both at training time and at inference time, the ResNet18 configurations would be the recommendation.

As the refinement algorithm is applied, improvements can be uniformly observed across all model configurations. By removing the bottom boundaries of the generated masks, the mean JSC is increased by 0.098, and the average DSC is increased by 0.026. The low variance of both similarity coefficients corroborates the observation that aggressive data augmentation is one of the main causes of irregularities in the generated masks (Table 4).

6. Conclusions

In this work, a series of experiments were run on Dynamic UNet architectures to determine their performance in chest cavity segmentation, achieving the best performance with a ResNet34 downsampling path, a self-attention layer, a blurring mechanism and a bottlenecked connection between the activation of the first block and the last block of the model. The average model performance without any post-processing achieved a DSC of 0.9359 and a JSC of 0.8874. Applying a simple algorithm aimed at correcting artefacts that were generated by aggressive data augmentation yielded an average DSC of 0.9612 and JSC 0.9789, achieving a new state of the art.

While the results are a significant step in the direction of clinical usability, future research should aim to expand the overall training datasets, as well as test the solution with different acquisition protocols and/or tasks.

Author Contributions: Conceptualization: M.B.; Methodology: M.B.; Software: M.B.; Validation: M.B., S.W. and W.B.; Investigation: M.B.; Data curation: S.W.; Writing: M.B., S.W. and W.B.; Visualization: M.B.; Reviewing and Editing: S.W. and W.B.; Supervision: T.-H.G. and W.B. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Review Board of HRA and HCRW (IRAS project ID 258617; latest amendment date: 18 December 2020).

Informed Consent Statement: Patient consent was waived due to the data collected being anonymized.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to ethical restrictions.

Conflicts of Interest: None top declare.

References

1. Bray, F.; Ferlay, J.; Soerjomataram, I.; Siegel, R.L.; Torre, L.A.; Jemal, A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **2018**, *68*, 394–424. [CrossRef]
2. Sung, H.; Ferlay, J.; Siegel, R.L.; Laversanne, M.; Soerjomataram, I.; Jemal, A.; Bray, F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J. Clin.* **2021**, *71*, 209–249. [CrossRef]
3. Ferlay, J.; Soerjomataram, I.; Dikshit, R.; Eser, S.; Mathers, C.; Rebelo, M.; Parkin, D.M.; Forman, D. Bray, Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. *Int. J. Cancer* **2014**, *136*, 359–386. [CrossRef]
4. Vinnicombe, S. How I report breast magnetic resonance imaging studies for breast cancer staging and screening. *Cancer Imaging* **2016**, *16*, 1–14. [CrossRef]
5. Cho, N.; Im, S.-A.; Park, I.-A.; Lee, K.-H.; Li, M.; Han, W.; Noh, D.-Y.; Moon, W.K. Breast cancer: Early prediction of response to neoadjuvant chemotherapy using parametric response maps for MR imaging. *Radiology* **2014**, *272*, 385–396. [CrossRef]
6. Zbontar, J.; Knoll, F.; Sriram, A.; Murrell, T.; Huang, Z.; Muckley, M.J.; Defazio, A.; Stern, R.; Johnson, P.; Bruno, M.; et al. fastMRI: An Open Dataset and Benchmarks for Accelerated MRI. *arXiv* **2018**, arXiv:1811.08839. [CrossRef]
7. Chang, Y.-C.; Huang, Y.-H.; Huang, C.-S.; Chang, P.-K.; Chen, J.-H.; Chang, R.-F. Magnetic Resonance Spectroscopy and Imaging Guidance in Molecular Medicine: Targeting and Monitoring of Choline and Glucose Metabolism in Cancer. *Magn. Reson. Imaging* **2012**, *30*, 312–322. [CrossRef] [PubMed]
8. Piantadosi, G.; Sansone, M.; Fusco, R.; Sansone, C. Multi-planar 3D breast segmentation in MRI via deep convolutional neural networks. *Artif. Intell. Med.* **2020**, *103*, 101781. [CrossRef] [PubMed]

9. Marrone, S.; Piantadosi, G.; Fusco, R.; Petrillo, A.; Sansone, M.; Sansone, C. Breast segmentation using Fuzzy C-Means and anatomical priors in DCE-MRI. In Proceedings of the 23rd IEEE International Conference on Pattern Recognition (ICPR), Cancun, Mexico, 4–8 December 2016; pp. 1472–1477.

10. Kayalibay, B.; Jensen, G.; van der Smagt, P. CNN-based Segmentation of Medical Imaging Data. *arXiv* **2017**, arXiv:1701.03056. [CrossRef]

11. Alshanbari, H.S.; Amin, S.; Shuttleworth, J.; Slman, K.A.; Muslam, S. Automatic Segmentation in Breast Cancer Using Watershed Algorithm. *Int. J. Biomed. Eng.* **2015**, *2*, 1–6.

12. Wang, L.; Platel, B.; Ivanovskaya, T.; Harz, M.; Hahn, H.K. Fully automatic breast segmentation in 3D breast MRI. In Proceedings of the 9th IEEE International Symposium on Biomedical Imaging (ISBI), Barcelona, Spain, 2–5 May 2012.

13. Vignati, A.; Giannini, V.; de Luca, M.; Morra, L.; Persano, D.; Carbonaro, L.A.; Bertotto, I.; Martincich, L.; Regge, D.; Bert, A.; et al. Performance of a Fully Automatic Lesion Detection System for Breast DCE-MRI. *J. Magn. Reson. Imaging* **2011**, *34*, 1341–1351. [CrossRef]

14. Gallego Ortiz, C.; Martel, A.L. Automatic atlas-based segmentation of the breast in MRI for 3D breast volume computation. *Med. Phys.* **2012**, *39*, 5835–5848. [CrossRef]

15. Gubern-Mérida, A.; Kallenberg, M.; Mann, R.M.; Martí, R.; Karssemeijer, N. Breast segmentation and density estimation in breast MRI: A fully automatic framework. *IEEE J. Biomed. Health Inform.* **2015**, *19*, 349–357. [CrossRef] [PubMed]

16. Khalvati, F.; Gallego-Ortiz, C.; Balasingham, S.; Martel, A.L. Automated Segmentation of Breast in 3D MR Images Using a Robust Atlas. *IEEE Trans. Med. Imaging* **2015**, *34*, 116–125. [CrossRef]

17. Reed, V.K.; Woodward, W.A.; Zhang, L.; Strom, E.A.; Perkins, G.H.; Tereffe, W.; Oh, J.L.; Yu, T.K.; Bedrosian, I.; Whitman, G.J.; et al. Automatic segmentation of whole breast using atlas approach and deformable image registration. *Int. J. Radiat. Oncol. Biol. Phys.* **2009**, *73*, 1493–1500. [CrossRef]

18. Fooladivanda, A.; Shokouhi, S.B.; Mosavi, M.R.; Ahmadinejad, N. Atlas-based automatic breast MRI segmentation using pectoral muscle and chest region model. In Proceedings of the 21th IEEE Iranian Conference on Biomedical Engineering (ICBME), Tehran, Iran, 26–28 November 2014.

19. Mustra, M.; Bozek, J. Breast border extraction and pectoral muscle detection using wavelet decomposition. In Proceedings of the IEEE EUROCON, St. Petersburg, Russia, 18–23 May 2009.

20. Wu, S.; Weinstein, S.P.; Conant, E.F.; Schnall, M.D.; Kontos, D. Automated chest wall line detection for whole-breast segmentation in sagittal breast MR images. *Med. Phys.* **2013**, *40*, 1–12. [CrossRef]

21. Cai, L.; Gao, J.; Zhao, D. A review of the application of deep learning in medical image classification and segmentation. *Ann. Transl. Med.* **2020**, *8*, 713. [CrossRef]

22. Dalmış, M.; Litjens, G.; Holland, K.; Setio, A.; Mann, R.; Karssemeijer, N.; Gubern-Mérida, A. Using deep learning to segment breast and fibroglandular tissue in MRI volumes. *Med. Phys.* **2017**, *44*, 533–546. [CrossRef]

23. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, 5–9 October 2015; pp. 234–241.

24. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention is All you Need. In Proceedings of the Advances in Neural Information Processing Systems 30 (NIPS 2017), Long Beach, CA, USA, 4–8 December 2017.

25. Iglovikov, V.I. Shvets, TernausNet. In *Computer-Aided Analysis of Gastrointestinal Videos*; Bernal, J., Histace, A., Eds.; Springer: Cham, Switzerland, 2021. [CrossRef]

26. Maitra, D.S.; Bhattacharya, U.; Parui, S.K. CNN based common approach to handwritten character recognition of multiple scripts. In Proceedings of the 13th IEEE International Conference on Document Analysis and Recognition (ICDAR), Tunis, Tunisia, 23–26 August 2015; pp. 1021–1025.

27. Raghu, M.; Zhang, C.; Kleinberg, J.; Bengio, S. Transfusion: Understanding Transfer Learning for Medical Imaging. In Proceedings of the Advances in Neural Information Processing Systems, Vancouver, BC, Canada, 8–14 December 2019.

28. Tan, C.; Sun, F.; Kong, T.; Zhang, W.; Yang, C.; Liu, C. A Survey on Deep Transfer Learning. In Proceedings of the Artificial Neural Networks and Machine Learning—ICANN 2018, Rhodes, Greece, 4–7 October 2018; pp. 270–279.

29. Serte, S.; Serener, A.; Al-Turjman, F. Deep learning in medical imaging: A brief review. *Trans. Emerg. Telecommun. Technol.* **2020**, *33*, e4080. [CrossRef]

30. Zhang, H.; Dauphin, Y.N.; Ma, T. Fixup Initialization: Residual Learning Without Normalization. *arXiv* **2019**, arXiv:1901.09321. [CrossRef]

31. He, K.; Zhang, X.; Ren, S.; Sun, J. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In Proceedings of the International Conference on Computer Vision ICCV, Santiago, Chile, 7–13 December 2015.

32. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. *Commun. ACM* **2017**, *60*, 84–90. [CrossRef]

33. Litjens, G.; Kooi, T.; Bejnordi, B.E.; Setio, A.A.A.; Ciompi, F.; Ghafoorian, M.; van der Laak, J.A.W.M.; van Ginneken, B.; Sánchez, C.I. A survey on deep learning in medical image analysis. *Med. Image Anal.* **2017**, *42*, 60–88. [CrossRef]

34. Razzak, M.I.; Naz, S.; Zaib, A. Deep Learning for Medical Image Processing: Overview, Challenges and the Future. *Classif. BioApps Lect. Notes Comput. Vis. Biomech.* **2017**, *26*, 323–350. [CrossRef]

35. Bahdanau, D.; Cho, K.; Bengio, Y. Neural Machine Translation by Jointly Learning to Align and Translate. In Proceedings of the 2nd International Conference on Learning Representations, Banff, AB, Canada, 14–16 April 2014.

36. Xu, T.; Zhang, P.; Huang, Q.; Zhang, H.; Gan, Z.; Huang, X.; He, X. AttnGAN: Fine-Grained Text to Image Generation with Attentional Generative Adversarial Networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017.

37. Yang, Z.; He, X.; Gao, J.; Deng, L.; Smola, A. Stacked attention networks for image question answering. *arXiv* **2015**, arXiv:1511.02274.

38. Gregor, K.; Danihelka, I.; Graves, A.; Rezende, D.J.; Wierstra, D. Draw: A recurrent neural network for image generation. *arXiv* **2015**, arXiv:1502.04623.

39. Guan, Q.; Huang, Y.; Zhong, Z.; Zheng, Z.; Zheng, L.; Yang, Y. Diagnose like a Radiologist: Attention Guided Convolutional Neural Network for Thorax Disease Classification. *arXiv* **2018**, arXiv:1801.09927. [CrossRef]

40. Zhang, H.; Goodfellow, I.; Metaxas, D.; Odena, A. Self-Attention Generative Adversarial Networks. In Proceedings of the 36th International Conference on Machine Learning, Long Beach, CA, USA, 9–15 June 2019; pp. 7354–7363.

41. Odena, A.; Dumoulin, V.; Olah, C. Deconvolution and Checkerboard Artifacts. *Distill* **2016**, *1*, e3. [CrossRef]

42. Mikolajczyk, A.; Grochowski, M. Data augmentation for improving deep learning in image classification problem. In Proceedings of the IEEE International Interdisciplinary PhD Workshop (IIPhDW), Świnoućcie, Poland, 9–12 May 2018; pp. 117–122.

43. Wong, S.C.; Gatt, A.; Stamatescu, V.; McDonnell, M.D. Understanding data augmentation for classification: When to warp? In Proceedings of the IEEE International Conference on Digital Image Computing: Techniques and Applications (DICTA), Gold Coast, Australia, 30 November–2 December 2016; pp. 1–6.

44. Kingma, D.P.; Ba, J. A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.

45. Smith, L.N. A disciplined approach to neural network hyper-parameters: Part 1—Learning rate, batch size, momentum, and weight decay. *arXiv* **2018**, arXiv:1803.09820.

46. Smith, L.N. No more pesky learning rate guessing games. *arXiv* **2015**, arXiv:1506.01186.