ABUSDet: A Novel 2.5D Deep Learning Model for Automated Breast Ultrasound Tumor Detection

Xudong Song[1], Xiaoyang Lu[1], Gengfa Fang[1*], Xiangjian He[2*], Xiaochen Fan[3], Le Cai[1], Wenjing Jia[1*] and Zumin Wang[4*]

[1*] University of Technology Sydney, Australia.
[2*] University of Nottingham Ningbo China.
[3] Tsinghua University, China.
[4*] Dalian University, China.

*Corresponding author(s). E-mail(s): Gengfa.Fang@uts.edu.au; Sean.He@nottingham.edu.cn; Wenjing.Jia@uts.edu.au; wangzumin@163.com;
Contributing authors: Xudong.Song@student.uts.edu.au; Xiaoyang.Lu@student.uts.edu.au; fanxiaochen33@gmail.com; 12635177@student.uts.edu.au;

Abstract

Automated Breast Ultrasound is a highly advanced breast tumor detection modality that produces hundreds of 2D slices in each scan. However, this large number of slices poses a significant burden for physicians to review. This paper proposes a novel 2.5D tumor detection model, named "ABUSDet," to assist physicians in automatically reviewing ABUS images and predicting the locations of breast tumors in images. At the core of this approach, a sequence of data blocks partitioned from a pre-processed 3D volume are fed to a 2.5D tumor detection model, which outputs a sequence of 2D tumor candidates. An aggregation module then clusters the 2D tumor candidates to produce the ultimate 3D coordinates of the tumors. To further improve the accuracy of the model, a novel mechanism for training deep learning models, called "Deliberate Training," is proposed. The proposed model is trained and tested on a dataset of 87 patients with

235 ABUS volumes. It achieves sensitivities of 77.94%, 75.49%, and 65.19% at FPs/volume of 3, 2, and 1, respectively. Compared with the 2D and 3D object detection models, the proposed ABUSDet model achieves the highest sensitivity with relatively low false-positive rates.

Keywords: Automated breast ultrasound (ABUS), breast cancer, 2.5D tumor detection, deliberate training.

1 Introduction

Breast cancer has overtaken lung cancer as the most commonly diagnosed cancer globally [1]. Early diagnosis and screening, in combination with appropriate treatment, can significantly reduce mortality rates over the long term [2]. According to a study that analyzed the World Health Organization Cancer Mortality Database from 1950 to 2016, and a review of the literature from 2010 to 2020, in countries that have achieved sustained reductions in breast cancer mortality rates, the majority of cases are diagnosed at an early stage [3]. This highlights the importance of early clinical diagnosis in improving breast cancer outcomes.

Mammography is the first-generation technology used for breast tumor detection. Having been extensively studied, mammography remains the primary tool for early diagnosis and screening of breast cancer. However, it is hampered because of the issue of contraindication [4]. Also, its sensitivity and specificity are, statistically, significantly lower than ultrasound in dense breasts and for young women [5]. As a result, Handheld Ultrasound (HHUS) and Magnetic Resonance Imaging (MRI) are frequently used as adjuncts to mammography for further evaluation of questionable findings [6]. While HHUS has limitations such as operator dependence and imperfect reproducibility, MRI has superior sensitivity but a high sensitivity to benign breast diseases [7]. The Automated Breast Ultrasound System (ABUS) [8] is a new ultrasound screening technology for breast tumor diagnosis that generates high-quality 3D volume data with high screening efficiency and low operational dependence. ABUS can produce hundreds of slices per scan with rich details. However, reviewing thousands of slices for each patient is time-consuming for physicians. Moreover, the varying size, shape, and texture of tumors for different patients makes a review even more challenging. Fig. 1 shows four ABUS slices with tumors from four patients, illustrating the differences in difficulty level among them. The lesion in the lower-left image is relatively easy to detect, while the others present more challenges for accurate prediction.

1.1 Related Work

To enhance the accuracy of lesion detection and alleviate the workload of physicians, researchers have proposed several computer-aided detection systems for ABUS tumor detection [9], [10], [11], [12], [13], [14], [15], [16], [17].

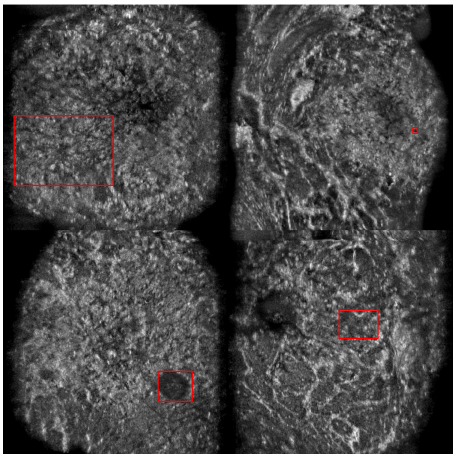


Fig. 1: Samples of lesions with different morphological patterns (Coronal View).

Machine Learning-based Approaches: In the field of medical imaging, the detection of tumors using machine learning techniques has seen significant advances in recent years. Prior to 2017, traditional machine learning approaches were commonly used, such as multi-scale blob detection based on Hessian analysis [9], a multi-stage system integrating a landmarks detection algorithm and a neural-network classifier [10], and a Multi-Dimensional Tumor Detection system using Topographic Watershed [11]. But these methods were limited by accuracy bottlenecks and high false-positive rates because of the need for professional expertise in selecting features.

Deep Learning-based Approaches: Deep learning-based solutions have gained significant attention in medical imaging applications, particularly in tumor detection, because they do not require hand-crafted features. Deep learning-based solutions have been found to outperform traditional machine learning methods. Various 3D object detection methods based on deep learning have been developed for detecting abnormalities. Chiang et al. [12] proposed a 3D CNN for estimating tumor probability by extracting the volume of interest (VOI) through sliding windows. Wang et al. proposed a 3D CNN and a dense deep supervision method to enhance detection sensitivity through multi-scale feature utilization [13]. Xiang et al. [16] proposed a 3D tumor diagnosis system by combining U-net, and a residual-capsule neural network. 3D object detection methods have the advantage of capturing detailed spatial information of the object being detected, but they have limitations such as the need for large amounts of computing power and limited annotated data.

Alternatively, 2D object detection methods based on deep learning, such as Faster R-CNN and YOLO, have also been widely used in medical imaging applications. Zhou et al. [14] proposed a 3D multi-view tumor detection method, which used a modified Faster R-CNN for 2D candidate extraction

and a 3D multi-view position analysis scheme to obtain final 3D bounding boxes. Li et al. [15] designed an improved YOLOv3 to detect tumors in two-dimensional slices, and then introduced a rescoring processing algorithm to reduce false positives. Zhang et al. [17] proposed a Bayesian YOLOv4 network based on Monte Carlo dropout (MC-Drop) to introduce uncertainty into the detection network. However, 2D methods often struggle to fully represent 3D objects, leading to decreased accuracy in object detection tasks.

1.2 Our Work

To combine the advantages of both 2D and 3D approaches and to improve sensitivity while minimizing false positives (FPs), we propose a novel 2.5D tumor detection method named ABUSDet. This method utilizes 3D depth information to enhance the accuracy and robustness of detection while leveraging the 2D method's optimized standardized model to improve computational efficiency.

Our proposed approach first partitions the input 3D volume data into 2.5D data blocks, which consists of multiple sequential slices. Compared to 2D data with only three channels, 2.5D data provides more information, while having fewer channels than the original 3D data with hundreds of channels. The depth information conveyed by continuous slices offers additional contextual cues that significantly enhance detection accuracy across various applications. By fully leveraging the benefits of both 2D and 3D approaches, our approach improves detection accuracy while reducing computational cost. Moreover, our 2.5D approach integrates a Channel Squeeze module with a 2D object detection module to condense multi-channel information into three channels, which are then used for the detection task. To further improve accuracy, we propose a novel mechanism for training the deep learning models called "Deliberate Training".

Finally, to evaluate the performance of our proposed ABUSDet method, we compare it with the state-of-the-art 2D and 3D methods using a set of performance metrics, including sensitivity, false positive rate, and computational cost. Experimental results demonstrate that our ABUSDet outperforms the 2D and 3D methods while maintaining a competitive computational cost, indicating its superiority in detecting breast tumors in ultrasound imaging.

The main contributions of this paper are as follows.

1) We propose a novel 2.5D tumor detection model by detecting tumors from 2.5D data and integrating a Channel Squeeze module to improve tumor detection performance. The 2.5D approach offers a combination of the fast and accurate benefits of 2D models and the robustness of 3D approaches, resulting in a reduction in training time and improved detection accuracy.

2) We propose a new mechanism for training deep learning models, i.e., deliberate training, which only trains the under-trained data and can improve the sensitivity of tumor detection by 5.0%.

3) This work provides a comprehensive solution in the field of breast tumor detection, integrating various concepts and techniques such as automated

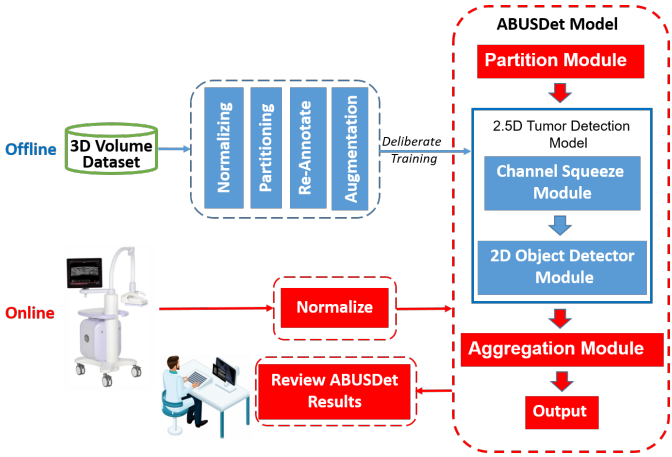


Fig. 2: The system architecture of the proposed ABUSDet.

breast ultrasound, breast cancer detection, a 2.5D tumor detection approach, and deliberate training.

4) We validate the effectiveness of the proposed ABUSDet through extensive experiments using our datasets, because there is no public ABUS dataset available. Experiment results show that our ABUSDet greatly outperforms the 2D, 2.5D, and 3D methods, demonstrating its superior effectiveness for localizing tumors in ABUS images.

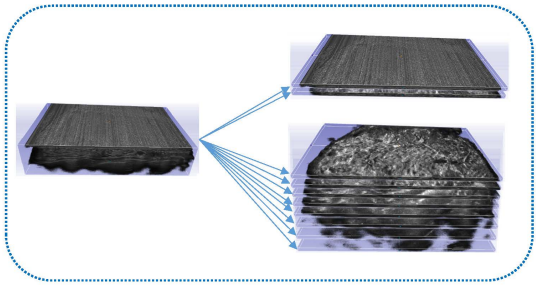
The rest of this paper is organized as follows. Sections 2 and 3 detail the proposed ABUSDet model and the new deliberate training mechanism. Section 4 presents the experiments and comparative results. Finally, Section 5 concludes the paper and suggests plans for future research.

2 Methodology

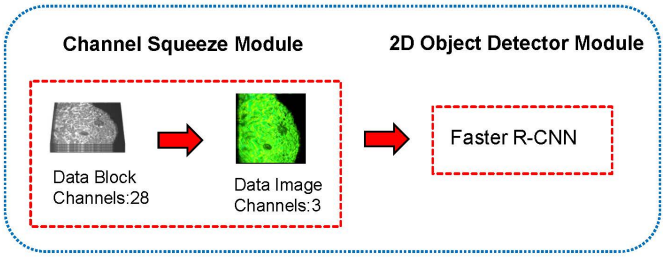
2.1 Overview

The ABUSDet system, as illustrated in Fig. 2, comprises offline and online stages. In the offline stage, the input training data undergo normalizing, partitioning, re-annotation, and augmentation. These datasets are then used to train the 2.5D tumor detection model using a newly developed "deliberate training" method. During the online stage, the input ABUS image is normalized before being processed by the proposed ABUSDet for tumor prediction. The system's output is then transmitted to the physician. The proposed deliberate training method focuses on the under-trained data, which presents a challenge in training in each epoch to further boost the sensitivity of tumor detection.

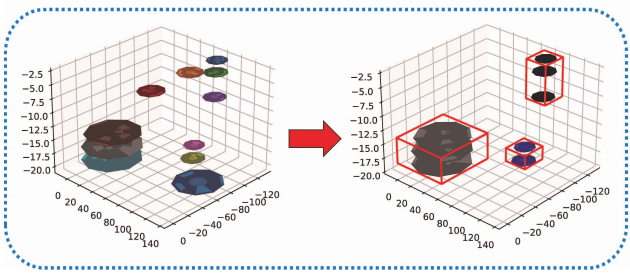
Our proposed ABUSDet is composed of three modules: a Data Partition module, a 2.5D Tumor Detection module, and an Aggregation module.



(a) The data partition module splits the 3D volume data into 2.5D data blocks (Coronal View).



(b) 2.5D tumor detection module



(c) The aggregation module. 2D tumor candidates (shown as colored disks) are clustered into 3D candidates (enclosed in the red boxes).

Fig. 3: The three modules of the proposed ABUSDet. (a) Data Partition Module. (b) Tumor Detection Module. (c) Aggregation Module.

First, the Data Partition module is used to partition 3D volume data into 2.5D data blocks, as illustrated in Fig. 3a. Subsequently, the 2.5D Tumor Detection module, shown in Fig. 3b, locates the lesion from the 2.5D data blocks. Finally, the Aggregation algorithm fuses the localization results of all blocks to produce the lesion location in the volume, as illustrated in Fig. 3c.

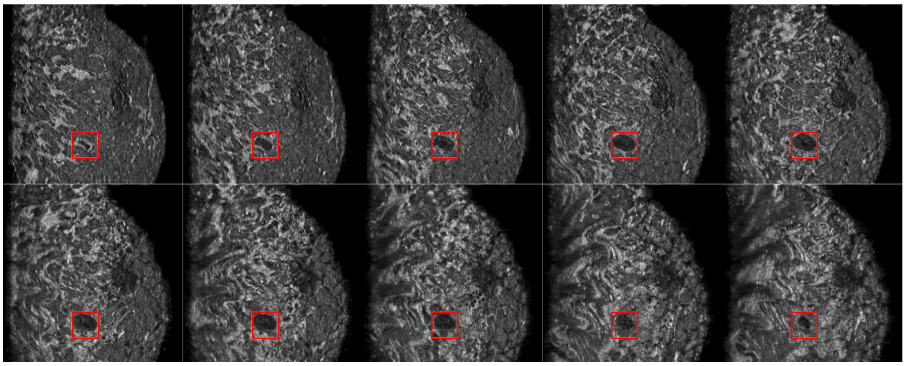


Fig. 4: A range of slices of data blocks showing one lesion.

2.2 Data Partition Module

Each DICOM (Digital Imaging and Communications in Medicine) volume consists of 330 slices (Axial view) arranged in a three-dimensional array. The dimensions of each frame vary due to differences in the acquired field of the view. To process the data, we first divide the 3D volume data into a series of multi-channel 2.5D data blocks. The shape of the 2.5D data block is determined by three parameters: splitting orientation, channel number, and stride. Splitting orientation determines the view obtained after data partition, with options including coronal, sagittal, and axial. The channel number specifies the number of channels contained within each block, and the stride determines the step size for cutting the 3D volume. In this study, the values of the channel number and stride were empirically set to 28 and 10, respectively. The splitting orientation is perpendicular to the coronal view. Fig. 3a shows how 3D volume is partitioned into 2.5D data blocks from a coronal view. For each volume, an experienced physician annotates each lesion. When the 2.5D data block is extracted during the offline stage, a 2D-formatted bounding box will be generated through a re-annotation procedure. Fig. 4 shows several 2.5D data blocks divided from a 3D volume that contains lesions.

Normalization: To ensure all the values in a volume are to the same scale, we normalize each 3D volume, first, before the partition module by

$$X' = \begin{cases} \dfrac{X - X_{min}}{X_{max} - X_{min}}, & X \leq X_{max} \\ 1, & X > X_{max} \end{cases}, \tag{1}$$

where X is the input 2.5D data block, X' is the output of the scaled feature, X_{min} is the minimum value among all pixels of the input data, and X_{max} is the maximum value of our data. Since most pixel values in the data fall between 0 and 200, we set the minimum X_{min} and maximum X_{max} values to 0 and 200, respectively. The normalized 3D volume data are then partitioned into multi-channel 2.5D data blocks at the initial stage of data processing.

Data Augmentation: The shortage of labeled datasets poses a significant challenge for medical imaging analysis. To address the problem of insufficient labeled data, data augmentation techniques are commonly employed. In the domain of 2D object detection, various image manipulation techniques such as rotation, scaling, and flipping have been used to expand the data set. Similarly, our 2.5D data blocks make use of these data augmentation techniques, which can be applied either online or offline. The offline approach generates data in advance and stores it in memory, while the online method processes data during training. To balance computational time and memory limitations, we use an online method for rotation, scaling, and flipping. Specifically, each data block has a 50% chance of being flipped when being loaded. Also we randomly rotate each block randomly by 30-degrees, except for 0, 90, 180, and 270 degrees, which are scaled to 75% of their original size. The following ABUSDet experiments are all based on online method augmented data.

2.3 2.5D Tumor Detection Module

Significant progress has been made in object detection from 2D images. Many successful models and methods have been proposed for medical imaging. To leverage these advances, we propose a modular 2.5D tumor detection model that combines a Channel Squeeze module and a 2D object detection model, resulting a faster and more accurate model that also considers the depth information. This model reduces the training time and improves training accuracy through transfer learning. We use Faster R-CNN as our 2D object detector module, as depicted in Fig. 3b.

After data partition and augmentation, we obtain multiple 2.5D data blocks from a 3D volume. A 2.5D data block is a multi-channel image that differs from traditional 2D pictures with one or three channels. To convert a multi-channel 2.5D data block to a 3-channel 2D image, we designed a Channel Squeeze module that is implemented using a convolutional neural network, as illustrated in Fig. 3b. This module can be integrated into the 2D object detection models to reduce the training time and improve training accuracy through transfer learning.

Note that our approach of using a channel squeeze module to squeeze multiple channels into three channels so as to adapt to a 2D object detection model is different from existing approaches with similar names. For example, He et al. [18] used a Squeeze-and-Excitation block to squeeze global spatial information into a channel descriptor.

Aggregation: The 2.5D tumor detection model outputs candidate sets with 2D coordinate positions, depth information, and tumor probability information. To obtain the final predicted 3D tumor location, we use DBSCAN [19], a density-based clustering non-parametric algorithm, to cluster the candidate sets. Fig. 3c illustrates the process.

When using DBSCAN, it is crucial to consider two important parameters: "min samples" and "eps". "Min samples" refers to the minimum number of samples required within a given neighborhood of a point in order for that point

to be classified as a core point. "Eps" refers to the maximum distance between two samples for them to be considered neighbors within a clustering algorithm. After experimentation, we found that setting "min samples" and "eps" in our dataset to 1 and 6mm, respectively, resulted in better accuracy compared to other values. Candidates with a tumor probability below the threshold of 0.2 are eliminated, and the highest probability value of each cluster subset is taken as the tumor probability.

Pre-training: Transfer learning is a crucial technique in deep learning that can help facilitate training. In the medical image field, transfer learning includes cross-domain models (pre-trained on natural images) and cross-modal models (pre-trained on medical images) models [20]. In the 2.5D tumor detection model of our ABUSDet, the 2D object detector model is Faster R-CNN with a backbone network ResNet-50 [21] pre-trained on ImageNet [22]. By using cross-domain transfer learning, only 24 epochs are needed to fine-tune the entire 2.5D tumor detection model.

3 Deliberate Training

Deliberate training is a new mechanism for training deep learning models. Training a model with the entire dataset is computationally inefficient. Therefore, to focus solely on the under-trained data in each training epoch, we propose a new deep learning training approach: deliberate training.

The inspiration for deliberate training comes from the deliberate practice proposed by psychologist Ericsson [23]. Ericsson identified a set of conditions that can significantly improve one's skill acquisition performance, which include: 1) a task with a well-defined goal, 2) motivation to improve, 3) provision with feedback, and 4) provision with ample opportunities for repetition and gradual refinement of performance.

We argue that the deliberate practice used to train the human brain can also be used to train neural network models in deep learning. The existing approaches used for training deep learning models already satisfy the above-mentioned criteria of deliberate practice. However, inspired by the fourth condition (ample opportunities for repetition), we propose to enhance the existing training process, which uses all the data in every epoch, by giving the under-trained data more opportunities to be used, repetitively, for training. This is achieved by ignoring the well-trained data in the current round.

Fig. 5 is the schematic diagram of our proposed deliberate training scheme for training the deep learning model. Algorithm 1 provides the pseudo-code of the algorithm. In Fig. 5, the number in red circles represents the important operation. Operation 1 is a Train Epoch corresponding to Lines 2 through 7 in Algorithm 1. Before each Training Epoch, Lines 2 through 3 in Algorithm 1 are executed. Prior to training, we define which epoch needs to be trained using the deliberate training method and save its indexes in the delibEpochs set. The initial value of delibEpochs is defined in Table 1. Line 2 in Algorithm 1

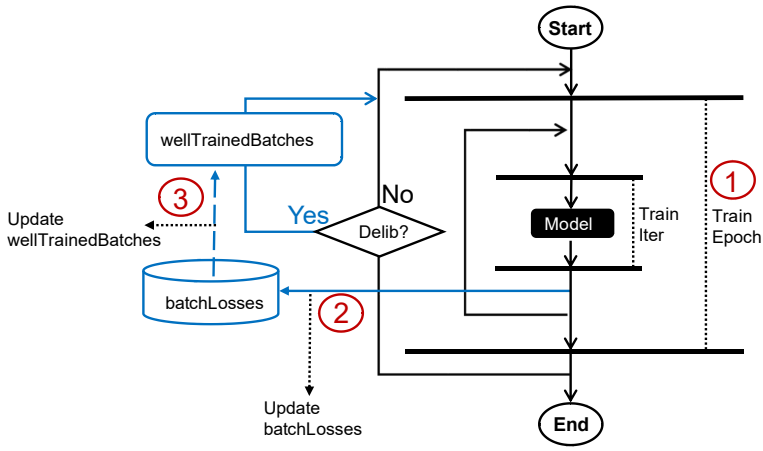


Fig. 5: The schematic diagram of our proposed deliberate training scheme for training the tumor detection model.

determines whether the current epoch requires deliberate training. If it is not required, we set the wellTrainedBatches to an empty set.

Operation 2 corresponds to Lines 5 to 6 in Algorithm 1 and is executed after each Train Iter in each Train Epoch. This operation computes the batch delibLoss and updates the batchLosses with it. Operation 3 corresponds to Lines 8 to 12 in Algorithm 1 and is performed after each Train Epoch. This operation evaluates the training performance of each batch based on the updated batchLosses. Next, some samples with lower training losses are excluded in this round of training. The median of batchLosses is used to decide which batchLosses are well-trained in the current epoch.

We compared three loss functions, namely, the positive hit number (phNum), the positive hit number with Non-maximum Suppression [24] (phNum+NMS), and the positive missed number (pmLoss). The phNum value indicates all the true-positive numbers of candidates in a batch. Based on the phNum, we apply NMS to the candidate of each image and obtain the positive hit numbers as the phNum+NMS. The larger the phNum or phNum+NMS in a batch, the more effectively the batch is trained. The pmLoss indicates the number of false negatives in a batch. The lower the pmLoss, the better the batch is trained.

Note that the deliberate training method is related to "hard sample mining" [25]. Hard sample mining is a widely used technique in machine learning, aiming at improving model performance by selectively training "hard" to categorize examples during model training. Therefore, hard sample mining focuses on identifying and learning from hard negative objects in an image. In contrast, the proposed deliberate training method focuses on hard batches, which allows for a more holistic optimization of difficult samples, while optimizing the

Algorithm 1 Deliberate Training

dataset: An ordered list that contains data batches for training.
numberBatch: Number of batches to be trained.
numberEpoch: The number of times to completely train all data.
delibEpochs: A set of epoch indexes that need to be executed in deliberate practice.
delibLoss: A loss calculated from a batch for deliberate training.
batchLosses: An ordered list to save delibLoss of each batch.
wellTrainedBatches: A set of batch indexes that have been trained relatively well.
median: Returns the median of the array elements.
trainGetLoss: A function used to Train and return delibLoss for a batch.

for $currentEpoch \in [0, numberEpoch]$ **do**
  if $currentEpoch \notin delibEpochs$ **then** $wellTrainedBatches \leftarrow \emptyset$
  end if
  for $currentBatch \in [0, numberBatch]$ **do**
    if $currentBatch \notin wellTrainedBatches$ **then** $batchLosses[currentBatch] \leftarrow trainGetLoss(dataset[currentBatch])$
    end if
  end for
  $wellTrainedBatches \leftarrow \emptyset$
  for $currentBatch \in [0, numberBatch]$ **do**
    if $batchLosses[currentBatch] < median(batchLosses)$ **then** $wellTrainedBatches \leftarrow wellTrainedBatches \cup \{currentBatch\}$
    end if
  end for
end for

training process itself, resulting in increased efficiency. Therefore, the deliberate training method can be seen as an extension of hard sample mining, with the additional advantage of optimizing the training process as a whole, rather than just focusing on individual hard samples.

4 Experiments and Discussion

To demonstrate the effectiveness of the proposed ABUSDet for localizing tumors in ABUS images, we conducted comparative experiments against 2D and 3D methods on real-world ABUS data collected from our partner hospital. In this section, we first provide details about the data. Then, we present the details of the ablation studies conducted, justifying the parameters used in our approach. Finally, qualitative and quantitative comparisons against 2D and 3D approaches are provided.

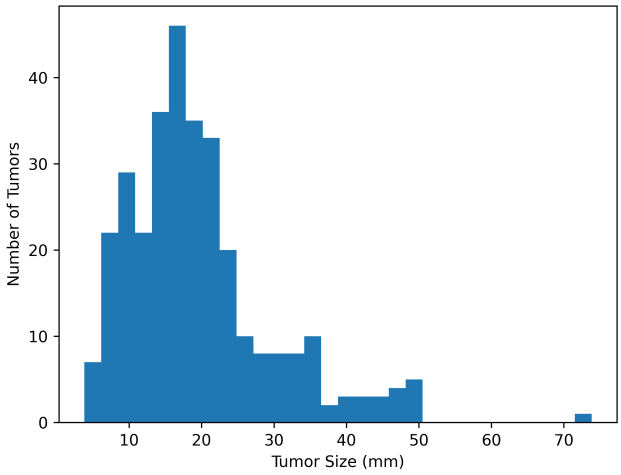


Fig. 6: Statistics of the size of the lesions.

4.1 Dataset, Evaluation Metrics and Implementation

4.1.1 Dataset

The ABUS dataset used in this study consists of 235 DICOM images from 87 patients between 23 to 65 years old, containing a total of 288 lesions. The data were collected between 2017 and 2019, using an Automated Breast Ultrasound Scanner (Invenia1.2) manufactured by GE Healthcare at the Traditional Chinese Medicine Hospital of Guangdong Province. The device generates 330 image slices with a Spacing Between Slices of 0.506mm and a Pixel Spacing of 0.082mm and 0.2mm along the posterior-anterior and left-right axes, respectively. The ABUS data used the Transducer Frequencies of 8.18MHz, 9MHz, or 10MHz.

Three regions are scanned in each breast for complete coverage of the patient, resulting in six DICOM images per patient, referred to as RAP, RLAT, RMED, LAP, LLAT, and LMED. However, in this study, only images that contain lesions are utilized to avoid unbalanced positive and negative samples. Fig. 6 shows the size distribution of the lesions (longest dimension). As shown in this figure, most lesions are smaller than 50mm with an average size of 19.6mm. The largest lesion is 73.8mm $\times$ 20.5mm $\times$ 51.1mm, and the smallest lesion is 2.9mm $\times$ 3.1mm $\times$ 3.8mm.

The patient data were randomly divided into three sets: Training, Validation, and Test Sets, consisting of 35, 17, and 35 patients, respectively. The Training, Validation, and Test Sets have 90, 44, and 101 volumes and a corresponding number of 102, 66, and 120 tumors, respectively.

4.1.2 Implementation

Table 1 presents the hyper-parameters used in ABUSDet, with their corresponding values at different stages of ABUSDet. In this work, Faster R-CNN is utilized as the 2D object detection module, and all the experiments were performed using MMDetection [26], an open-source object detection toolbox based on PyTorch [27]. The experiments were carried out on a computer equipped with an Intel(R) Xeon(R) Gold 6126 CPU with a 2.60GHz processor and a Tesla V100 32GB graphics card.

Table 1: The hyper-parameters of the proposed architecture.

Stage	Hyper-parameter	Value
Model	2D object detection model	Faster R-CNN ResNet-50
	Data Block Channels	28
Training	Deliberate Practice Loss	"phNum+NMS"
	Epochs	24
	Batch size	4
	delibEpochs	[1,5,10,15,20,21,22,23]
Testing	Confidence	0.2
	Stride	10
Aggregation	"min samples"	1
	"eps"	6mm
Optimizer	Type	"SGD"
	Learning Rate	0.005
	Momentum	0.9
	Weight Decay	0.0001
Learning Rate Policy	Type	"CosineAnnealing"
	Warmup	"exp"
	Warmup Epochs	8
	Warmup Ratio	0.1
	Min Lr	0.0001

4.1.3 Evaluation Metrics

The performance of the proposed method was evaluated using the free-response receiver operating characteristics (FROC) curves [28] and the area under the curve (AUC) [29]. The y-axis of the FROC plot represents sensitivity, while the x-axis represents the average value of FPs/volume. A predicted bounding box is considered a true positive if its centroid falls within the ground truth bounding box and the Euclidean distance between the centroids of the predicted and ground truth bounding boxes is less than 15mm.

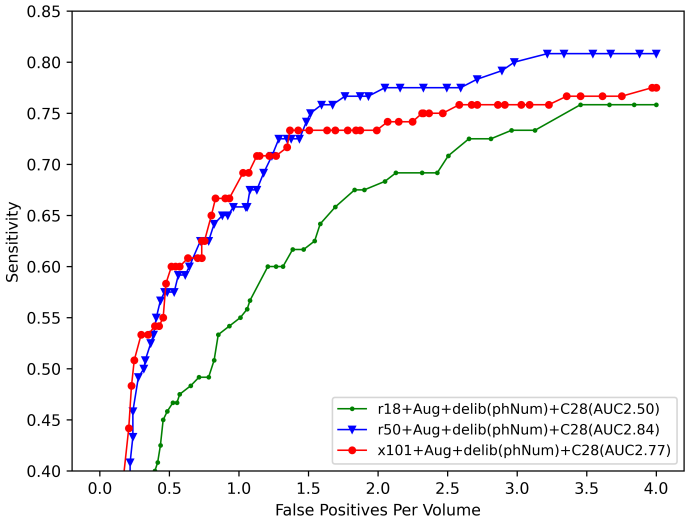


Fig. 7: Performance comparison of different Faster R-CNN models.

4.2 Ablation Studies

Ablation experiments were performed to evaluate the impact of the key modules and parameters in our proposed 2.5D tumor detection model, including the number of layers, Channel number, stride, and deliberate losses.

Impact of Different Numbers of Layers in the Backbone Network: Fig. 7 shows the results of the Faster R-CNN model with different layer numbers. The 50-layer Faster R-CNN model outperformed the 101-layer model, achieving the best results with the same training data and rounds. Theoretically, the 101-layer model needs more data for training.

Impact of the Deliberate Losses: Fig. 8 shows the results of regular training (without deliberate loss) and deliberate training with different deliberate losses, i.e., pmLoss, phNum, or phNum+NMS. The experimental results verify that the sensitivity is improved from 78.33% to 83.33%, and the AUC increased from 2.65 to 3.06 after using deliberate training with phNum+NMS. Moreover, it can be seen from the curve in Fig. 8 that deliberate training results are better than ordinary training methods for the sensitivity corresponding to any FPs/volume value. Note that the below control experiments are conducted based on phNum as a loss function instead of the optimal phNum+NMS.

Impact of the Channel Number: Fig. 9 shows the results of 2.5D tumor detection with different channel numbers. The results indicate that with a fixed stride, sensitivity and AUC increase with the increase of channel numbers, and it reaches the optimal value when the number of channels is 28. Detection using multiple channels outperformed 2D object detection. However, a channel number of 35 missed small and thin tumors, resulting in lower sensitivity.

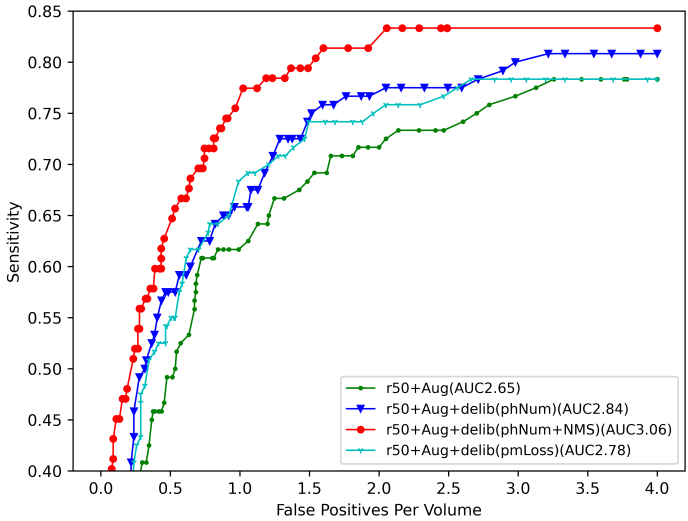


Fig. 8: Performance comparison of using different deliberate losses and without using a deliberate loss.

Table 2: Prediction time per volume with different strides in the ABUSDet model.

Stride	3	6	10	14	21	28
Time	35.5s	17.7s	11.0s	6.6s	5.7s	**3.8s**

Impact of Different Strides: Fig. 10 shows the results of different strides when deliberate training is utilized and channel numbers are set to 28. The maximum sensitivity was achieved with a stride of 10, whereas the highest AUC was obtained with a stride of 14. The larger the stride is, the faster is the calculation of a volume. When channels are set to 28, the test results under different strides show similar performance. The execution time depends on how the volume was partitioned. When the channel number is fixed, the larger the stride is, the lower the execution time. Table 2 provides the execution time of these methods.

Visualization of Detection Results: Based on the above experiments, we determined the optimal parameters of the ABUSDet model (see Table 1). Fig. 11 displays some samples of the detected tumors obtained on the test dataset using the optimal parameters. The figure shows the results of lesion detection in 2.5D data blocks using the ABUSDet model. Here, the images in the first row are samples with accurate tumor detection. Note the slices in the second row, where the nipple was wrongly recognized as a tumor, while ground truth tumors are detected. The ABUS image of the nipple can be easily confused with the lesion, even if recognized by experts. Therefore, insufficient

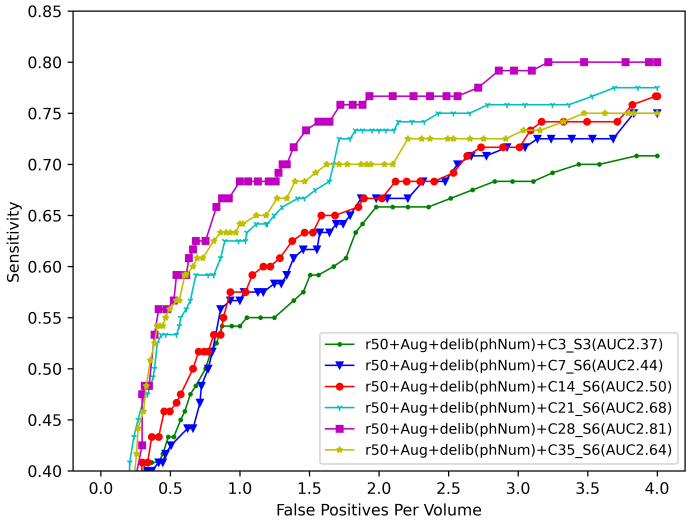


Fig. 9: Performance comparison of different channel numbers.

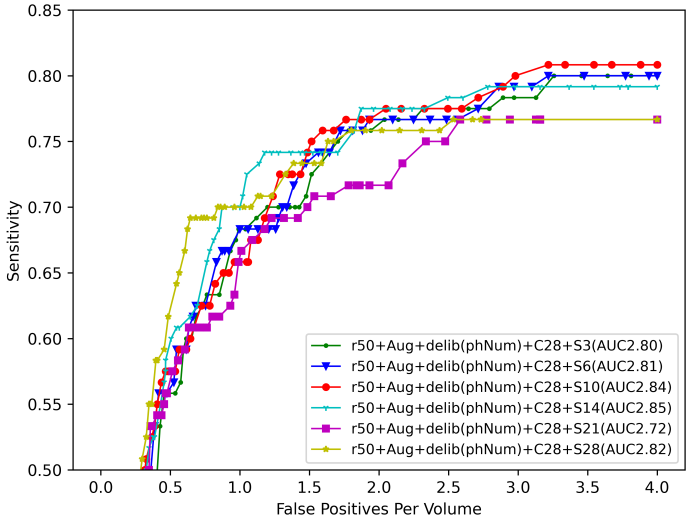


Fig. 10: Performance comparison obtained with different strides.

training data can cause the model to produce many FPs. The samples in the third row show the detected tumors and some non-nipple FPs.

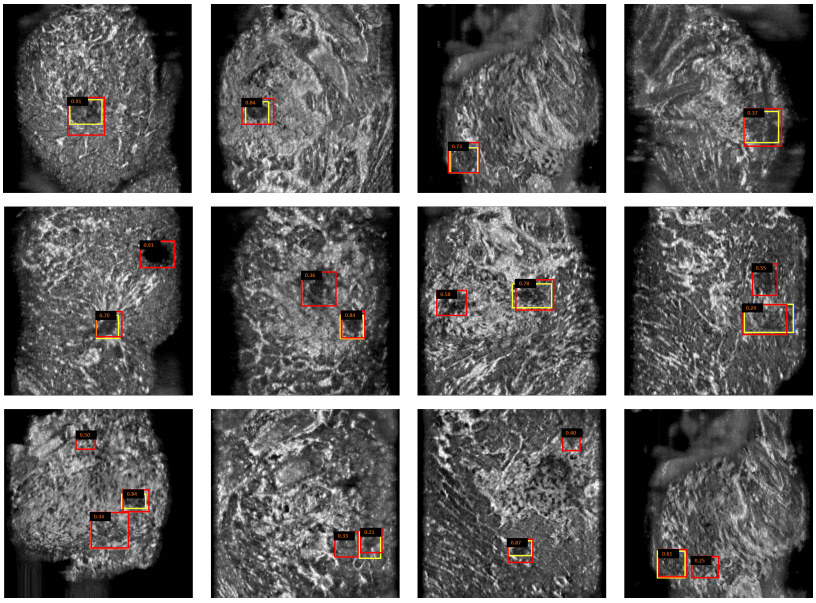


Fig. 11: Examples of tumor detection results on 2.5D data block of the Test set using the proposed ABUSDet approach. The yellow boxes indicate the ground-truth location of tumors, and the red boxes indicate the predicted location.

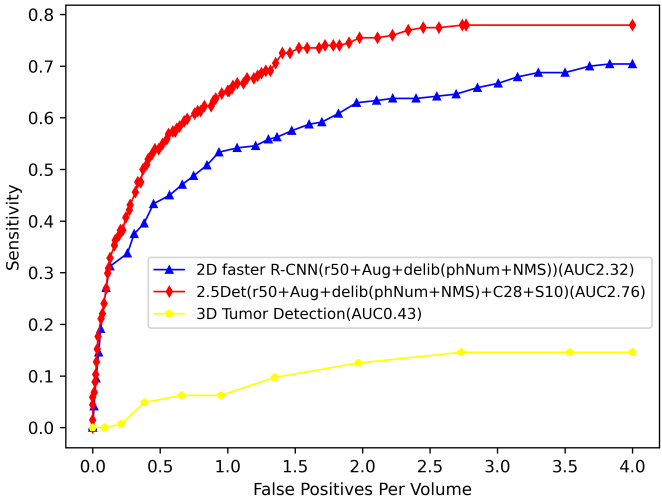


Fig. 12: Performance comparison of 2D, 3D, and 2.5D methods.

Table 3: Sensitivity and FPs/volumes of 2D, 3D and our proposed 2.5D approaches.

FPs/volume	Sensitivity(%)		
	2D Faster R-CNN	3D Tumor Detection [12]	ABUSDet
0.5	44.02	5.45	**54.31**
1.0	53.73	6.67	**65.19**
1.5	57.74	10.39	**73.12**
2.0	63.04	12.58	**75.49**
2.5	64.03	13.95	**77.45**
3.0	66.63	14.58	**77.94**
3.5	68.75	14.58	**77.94**
4.0	70.41	14.58	**77.94**
Execution time per volume	**8.5s**	67.0s	11.0s

4.3 Comparison with 2D and 3D Approaches

To compare the performance of 2D, 2.5D, and 3D object detection methods on ABUS data, we implemented a 2D object detection method based on the Faster R-CNN model and the 3D tumor detection method proposed in [12]. The 3D method generates candidate regions from preprocessed ABUS images and utilizes prioritized candidate aggregation to combine top-ranked regions into a final tumor probability map. We evaluated the performance of these methods on our dataset using 2-fold cross-validation. For a fair comparison, the training and clustering parameters of the 2D Faster R-CNN model are consistent with those of the proposed ABUSDet model. At the same time, the same evaluation metrics are adopted in these methods.

Fig. 12 compares the performance of the three approaches, where each curve represents the average accuracy of each fold under different false positive rates. As the result shows, the proposed ABUSDet has achieved the highest sensitivity, while the results of 3D tumor detection are very poor. We carefully investigate the reasons for the unsatisfactory performance of the 3D method. First of all, in our dataset, blurred tumors (such as those shown in the top-left and bottom-right corners in Fig. 1) which account for 55.8%, and small targets (such as the one in the top-right corner) which comprise 11.5% , are the most frequently observed lesions. In contrast, only 32.7% of tumors are sufficiently clear to quickly identify (such as the one shown in the bottom-left image). Second, the sensitivity of the 3D method can reach as high as 72% when the maximum number of allowed false positives per volume is set to eight, and when the criterion is lowered to "A predicted box will be treated as a true positive if the distance between the center of the predicted box and the center of the ground truth box is less than 25mm". Based on these two observations, it can be concluded that the 3D method may have limited sensitivity in detecting blurred and tiny tumors. Specifically, our dataset revealed that not all tumors were of a size suitable for detection using the 3D method's sliding window.

Moreover, given the time cost, it was not feasible to scan the 3D sliding window in a small stride to achieve fine detection. Consequently, the size and stride of the 3D sliding window may restrict the accuracy of tumor localization. Table 3 quantitatively compares the sensitivity and per-volume execution time of the three different methods. The execution times of the proposed ABUSDet and 2D Faster R-CNN's are considerably shorter than that of the 3D approaches (11 and 8.5 seconds *vs* 67 seconds).

5 Conclusion

This paper has proposed a novel 2.5D tumor detection model ABUSDet, which effectively detects tumors from 3D volume ABUS data without involving high computation costs. By integrating a Channel Squeeze module into the 2D object detector module, ABUSDet can deal with multi-channel data blocks and utilize the trained weights from other 2D object detection models. Finally, in the aggregation module, the 3D coordinates of each tumor candidate are recalculated according to their 2D coordinates and depth information, and then the 3D coordinates are clustered by the DBSCAN algorithm. In addition, we have devised a deliberate training mechanism in the training process. The experimental results demonstrate that the proposed 2.5D tumor detection with multiple channels is superior to 2D and 3D object detection methods. The results of the deliberate training mechanism outperform the traditional training process.

In summary, the proposed tumor detection model, ABUSDet, is a reliable and efficient assistant tool for reviewing breast tumors in ABUS volume. In the future, a transformer-based detection approach will be adopted as the primary detection structure; Additionally, unsupervised learning will be introduced for feature extraction, enabling large-scale unlabeled data to contribute to the detection performance.

Acknowledgments. We want to thank the Traditional Chinese Medicine Hospital of Guangdong Province for providing data for this study.

Declarations

Funding. This work is supported by the National Natural Science Foundation Grant (61976037) of China. It is also partly supported by Yongjiang Technology Innovation Project (2022A-097-G).

Competing Interests. The authors declare that they have no competing interests.

Data Availability Statements. Due to ethical concerns, supporting data cannot be made openly available.

Author's Contributions. Xudong Song: Investigation, data curation, writing the manuscript, and conducting the experiments. Xiaoyang Lu: Replicated

Ref [8]. Gengfa Fang: Investigation, data curation, project administration, supervision, writing, review, commentary, and revision. Xiangjian He: Funding acquisition, supervision, writing, review, commentary, and revision. Xiaochen Fan: Data analysis, review, commentary. Le cai: Data analysis, review, commentary. Wenjing Jia: Writing, review, commentary, and revision. Zumin Wang: Writing, review, commentary, and revision.

References

[1] 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: a cancer journal for clinicians **71**(3), 209–249 (2021)

[2] Mridha, M.F., Hamid, M.A., Monowar, M.M., Keya, A.J., Ohi, A.Q., Islam, M.R., Kim, J.-M.: A comprehensive survey on deep-learning-based breast cancer diagnosis. Cancers **13**(23), 6116 (2021)

[3] Duggan, C., Trapani, D., Ilbawi, A.M., Fidarova, E., Laversanne, M., Curigliano, G., Bray, F., Anderson, B.O.: National health system characteristics, breast cancer stage at diagnosis, and breast cancer mortality: a population-based analysis. The Lancet Oncology **22**(11), 1632–1642 (2021)

[4] Zamora, K., Allen, E., Hermecz, B.: Contrast mammography in clinical practice: Current uses and potential diagnostic dilemmas. Clinical Imaging **71**, 126–135 (2021)

[5] Zhang, Z., Wang, W., Wang, X., Yu, X., Zhu, Y., Zhan, H., Chen, Z., Li, B., Huang, J.: Breast-specific gamma imaging or ultrasonography as adjunct imaging diagnostics in women with mammographically dense breasts. European Radiology **30**(11), 6062–6071 (2020)

[6] Kuhl, C.: Mri of breast tumors. European radiology **10**(1), 46–58 (2000)

[7] Mann, R.M., Balleyguier, C., Baltzer, P.A., Bick, U., Colin, C., Cornford, E., Evans, A., Fallenberg, E., Forrai, G., Fuchsjäger, M.H., *et al.*: Breast mri: Eusobi recommendations for women's information. European radiology **25**(12), 3669–3678 (2015)

[8] Łuczyńska, E., Pawlak, M., Popiela, T., Rudnicki, W.: The role of abus in the diagnosis of breast cancer. Journal of Ultrasonography **22**(89), 76–85 (2022)

[9] Moon, W.K., Shen, Y.-W., Bae, M.S., Huang, C.-S., Chen, J.-H., Chang, R.-F.: Computer-aided tumor detection based on multi-scale blob detection algorithm in automated breast ultrasound images. IEEE Transactions

on Medical Imaging **32**(7), 1191–1200 (2013). https://doi.org/10.1109/TMI.2012.2230403

[10] Tan, T., Platel, B., Mus, R., Tabár, L., Mann, R.M., Karssemeijer, N.: Computer-aided detection of cancer in automated 3-d breast ultrasound. IEEE Transactions on Medical Imaging **32**(9), 1698–1706 (2013). https://doi.org/10.1109/TMI.2013.2263389

[11] Lo, C.-M., Chen, R.-T., Chang, Y.-C., Yang, Y.-W., Hung, M.-J., Huang, C.-S., Chang, R.-F.: Multi-dimensional tumor detection in automated whole breast ultrasound using topographic watershed. IEEE Transactions on Medical Imaging **33**(7), 1503–1511 (2014). https://doi.org/10.1109/TMI.2014.2315206

[12] Chiang, T.-C., Huang, Y.-S., Chen, R.-T., Huang, C.-S., Chang, R.-F.: Tumor detection in automated breast ultrasound using 3-d cnn and prioritized candidate aggregation. IEEE Transactions on Medical Imaging **38**(1), 240–249 (2019). https://doi.org/10.1109/TMI.2018.2860257

[13] Wang, Y., Wang, N., Xu, M., Yu, J., Qin, C., Luo, X., Yang, X., Wang, T., Li, A., Ni, D.: Deeply-supervised networks with threshold loss for cancer detection in automated breast ultrasound. IEEE Transactions on Medical Imaging **39**(4), 866–876 (2020). https://doi.org/10.1109/TMI.2019.2936500

[14] Zhou, Y., Chen, H., Li, Y., Wang, S., Cheng, L., Li, J.: 3d multi-view tumor detection in automated whole breast ultrasound using deep convolutional neural network. Expert Systems with Applications **168**, 114410 (2021). https://doi.org/10.1016/j.eswa.2020.114410

[15] Li, Y., Wu, W., Chen, H., Cheng, L., Wang, S.: 3d tumor detection in automated breast ultrasound using deep convolutional neural network. Medical Physics **47**(11), 5669–5680 (2020)

[16] Xiang, H., Huang, Y.-S., Lee, C.-H., Chien, T.-Y.C., Lee, C.-K., Liu, L., Li, A., Lin, X., Chang, R.-F.: 3-d res-capsnet convolutional neural network on automated breast ultrasound tumor diagnosis. European Journal of Radiology **138**, 109608 (2021)

[17] Zhang, Z., Li, Y., Wu, W., Chen, H., Cheng, L., Wang, S.: Tumor detection using deep learning method in automated breast ultrasound. Biomedical Signal Processing and Control **68**, 102677 (2021). https://doi.org/10.1016/j.bspc.2021.102677

[18] Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7132–7141 (2018)

[19] Ester, M., Kriegel, H.-P., Sander, J., Xu, X., *et al.*: A density-based algorithm for discovering clusters in large spatial databases with noise. In: Kdd, vol. 96, pp. 226–231 (1996)

[20] Ayana, G., Dese, K., Choe, S.-w.: Transfer learning in breast cancer diagnoses via ultrasound imaging. Cancers **13**(4), 738 (2021)

[21] He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)

[22] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp. 248–255 (2009). https://doi.org/10.1109/CVPR.2009.5206848

[23] Anders Ericsson, K.: Deliberate practice and acquisition of expert performance: a general overview. Academic emergency medicine **15**(11), 988–994 (2008)

[24] Neubeck, A., Van Gool, L.: Efficient non-maximum suppression. In: 18th International Conference on Pattern Recognition (ICPR'06), vol. 3, pp. 850–855 (2006). IEEE

[25] Felzenszwalb, P.F., Girshick, R.B., McAllester, D., Ramanan, D.: Object detection with discriminatively trained part-based models. IEEE transactions on pattern analysis and machine intelligence **32**(9), 1627–1645 (2010)

[26] Chen, K., Wang, J., Pang, J., Cao, Y., Xiong, Y., Li, X., Sun, S., Feng, W., Liu, Z., Xu, J., et al.: Mmdetection: Open mmlab detection toolbox and benchmark. arXiv preprint arXiv:1906.07155 (2019)

[27] Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al.: Pytorch: An imperative style, high-performance deep learning library. Advances in neural information processing systems **32** (2019)

[28] Chakraborty, D.P.: Maximum likelihood analysis of free-response receiver operating characteristic (froc) data. Medical physics **16**(4), 561–568 (1989)

[29] Lobo, J.M., Jiménez-Valverde, A., Real, R.: Auc: a misleading measure of the performance of predictive distribution models. Global ecology and Biogeography **17**(2), 145–151 (2008)