scientific reports

Check for updates

OPEN

3-D breast nodule detection on automated breast ultrasound using faster region-based convolutional neural networks and U-Net

Kangrok Oh[1], Si Eun Lee[2] & Eun-Kyung Kim[2]✉

Mammography is currently the most commonly used modality for breast cancer screening. However, its sensitivity is relatively low in women with dense breasts. Dense breast tissues show a relatively high rate of interval cancers and are at high risk for developing breast cancer. As a supplemental screening tool, ultrasonography is a widely adopted imaging modality to standard mammography, especially for dense breasts. Lately, automated breast ultrasound imaging has gained attention due to its advantages over hand-held ultrasound imaging. However, automated breast ultrasound imaging requires considerable time and effort for reading because of the lengthy data. Hence, developing a computer-aided nodule detection system for automated breast ultrasound is invaluable and impactful practically. This study proposes a three-dimensional breast nodule detection system based on a simple two-dimensional deep-learning model exploiting automated breast ultrasound. Additionally, we provide several postprocessing steps to reduce false positives. In our experiments using the in-house automated breast ultrasound datasets, a sensitivity of 93.65% with 8.6 false positives is achieved on unseen test data at best.

Breast ultrasound is a widely employed imaging modality for cancer detection and diagnosis. Specifically, breast ultrasound has been used as an attractive supplementary imaging modality to mammography for dense breasts[1]. Conventional two-dimensional (2-D) hand-held ultrasound (HHUS) imaging has limitations such as high operator dependence, long examination time, and poor reproducibility[2]. Recently automated breast ultrasound (ABUS) imaging has been an emerging modality resolving the limitations of HHUS[3–11]. ABUS imaging system produces volumetric images which enable new diagnostic information, i.e., coronal plane reconstruction[12]. However, ABUS imaging requires a long interpretation time due to the vast amount of images for the diagnosis. Hence, developing computer-aided detection (CADe) systems for ABUS is critical and valuable in the field.

Breast nodule detection has been investigated thoroughly for HHUS images, where the applied techniques are effective for those using ABUS images. Yap et al.[13] proposed to adopt patch-based LeNet[14], U-Net[15], and a transfer learning approach based on pre-trained fully connected networks (FCN-AlexNet[16,17]) for breast lesion detection. Cao et al.[18] presented an experimental study comparing the fast region-based convolutional neural networks (Fast R-CNN)[19], Faster R-CNN[20], you only look once (YOLO)[21], YOLO version3 (YOLOv3)[22], and single shot multibox detector (SSD)[23] for breast lesion detection. Similarly, Yap et al.[24] proposed a transfer learning approach based on the Faster R-CNN, and Li et al.[25] proposed a deep learning model based on unsupervised region proposal and the ResNet-50[26] architecture for breast lesion detection. These state-of-the-art studies regarding breast nodule detection using HHUS images adopted advanced deep-learning models to guarantee a certain level of detection performance and generalization capability.

Those pioneering studies[6,27] regarding breast nodule detection for ABUS images utilized a combination of image pre-processing, feature extraction, and classification techniques such as multi-scale blob detection[6], and watershed transform[28]. These studies tend to show low generalization capability, while various deep-learning

[1]Research Institute of Radiological Science and Center for Clinical Imaging Data Science, Yonsei University College of Medicine, Seoul 03722, Republic of Korea. [2]Department of Radiology, Yongin Severance Hospital, Yonsei University College of Medicine, 363, Dongbaekjukjeon-daero, Giheung-gu, Yongin, Gyeonggi-do 16995, Republic of Korea. ✉email: ekkim@yuhs.ac

models were adopted and proposed in recent studies. For 2-D breast lesion detection for ABUS images, Zhang et al.[29] introduced a Bayesian YOLO version4 (YOLOv4) architecture by applying the uncertainty scheme into the YOLOv4[30] model. However, this study is limited to 2-D lesion detection using ABUS images. To detect three-dimensional (3-D) tumors in ABUS images, Moon et al.[31] and Chiang et al.[32] proposed systems based on 3-D convolutional neural networks (CNN). Lei et al.[33,34] presented a 3-D patch-based Mask scoring R-CNN for 3-D breast tumor segmentation, while the study focused on solving the segmentation problem. Because the aforementioned 3-D breast nodule detection systems were constructed using 3-D patch-based CNN architectures, they inevitably operate in a sliding window scheme at the testing phase to locate suspicious breast lesions. Locating 3-D breast nodules using 2-D was investigated in several studies. Zhang et al.[35] adopted YOLO version5 (YOLOv5)[36] as a detector and 3-D ResNet as a classifier to reduce false positives (FPs). A convolutional BiLSTM network was proposed by Malekmohammadi et al.[37], where the approximate location of 3-D breast lesions was provided as the format of a heat map. Zhou et al.[38] investigated a fusion of multi-view 2-D breast tumor detection outcomes by modifying Faster R-CNN architecture for 3-D breast tumor detection.

In this study, we aim to detect 3-D breast nodules using a simple 2-D deep learning methodology. To accomplish this goal, we employ the Faster R-CNN model due to the advantages of sharing convolutional features for region proposal and object detection and avoiding excessive search by a sliding window strategy. Subsequently, we propose a couple of additional postprocessing steps for removing false positives (FPs) to improve the detection performance. In addition, transverse, coronal, and sagittal plane images are utilized as input images because it is an advantage of ABUS volumetric imaging.

Methods

In this study, we propose an ABUS-based 3-D breast nodule detection system. The faster region-based convolutional neural networks (Faster R-CNN) model[20], which has been frequently employed for breast nodule detection on HHUS and ABUS images[18,38], is adopted to locate suspicious breast nodules from 2-D images in transverse, coronal, and sagittal planes. The 2-D FPs and redundant bounding boxes are eliminated based on a thresholding operation on output confidence values and clustering technique. Reduced 2-D bounding boxes extracted from 2-D image slices are then converted to 3-D cuboids based on a set of computations using row, column, and slice coordinates. Subsequently, redundant and FP 3-D cuboids are excluded based on U-Net segmentation model[15] and coordinates processing. Consequently, breast nodules in the form of 3-D cuboids are detected via the aforementioned 2-D deep learning methodology and postprocessing. Figure 1 shows an overview of the proposed 3-D breast nodule detection system.

Automated breast ultrasound data acquisition and preprocessing

We acquired ABUS images from Yongin Severance Hospital, Yonsei University College of Medicine, South Korea, to evaluate breast nodule detection performance. The Yongin Severance Hospital Institutional Review Board approved this study (IRB Number: 9-2021-0107). All research was performed under relevant guidelines and regulations. Informed consent was waived by the Institutional Review Board because of the retrospective
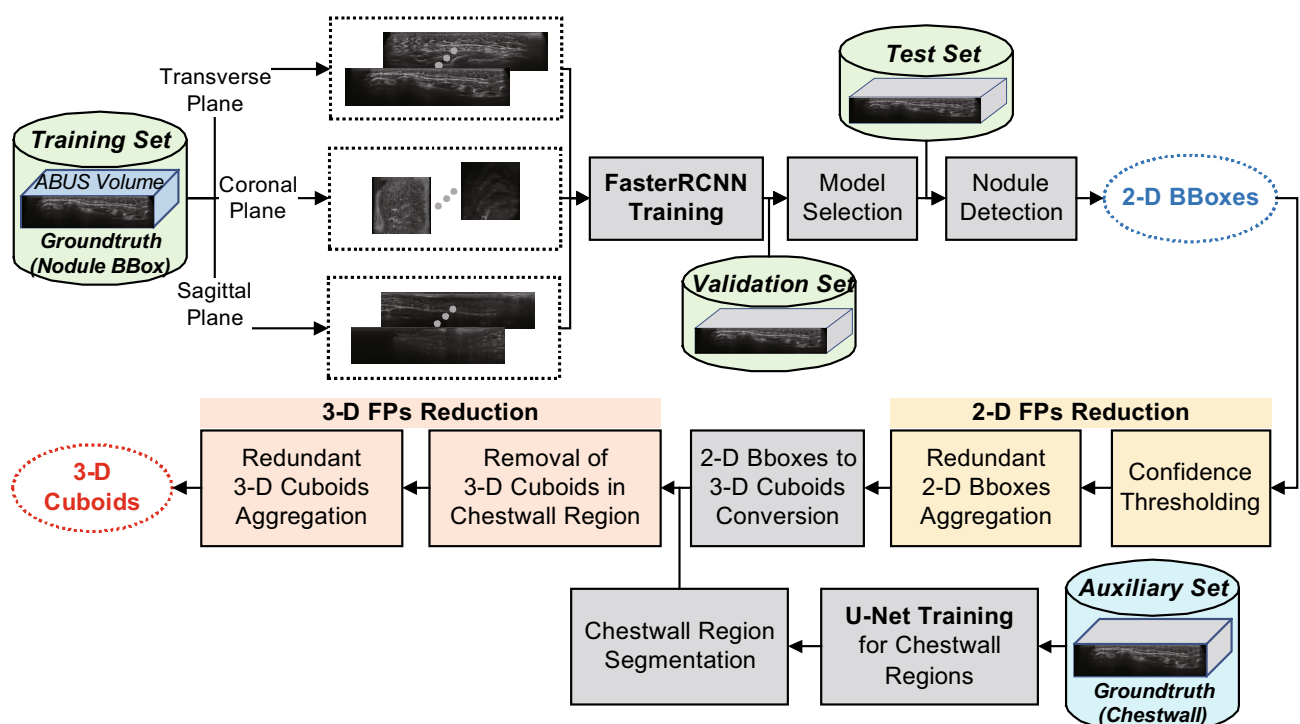

Figure 1. An overview of the proposed 3-D breast nodule detection system.

nature of the study, and the analysis used anonymous clinical data. The capturing device is Invenia ABUS 2.0 from GE Healthcare for entire ABUS images. The captured ABUS scans have the form of stacks of 2-D images in the transverse plane. The 2-D image in the transverse plane has the size of 546×843, where the number of transverse slices is around 350. Pixel spacing values in the longitudinal, lateral, and vertical axes are 0.073, 0.2, and 0.4757 millimeters. Resampling is performed to reformat the ABUS images as isotropic scans with 0.2 mm voxel spacing. Consequently, the reformatted ABUS scans have the size of $198 \times 842 \times 830$ pixels. Loading DICOM files and reformatting them to a stack of gray-scale images with isotropic voxel spacing was performed using MATLAB R2022a[39]. Specifically, the 'dicomread' and 'interp3' functions are the main components for processing. For the PyTorch implementation of the Faster R-CNN model for gray-scale images, converting the gray-scale images to RGB images by repeatably concatenating the gray-scale images according to Yap et al.[24], and rescaling the intensity values within [0, 1] was performed.

The in-house ABUS dataset A acquired from April 2020 to January 2021 for training and testing consists of 308 scans/passes from 103 female patients. It contains 240 scans with benign-appearing nodules (Breast Imaging Reporting and Data System (BI-RADS)[40] categories 2 and 3) from 74 patients and 68 scans with suspicious nodules (BI-RADS categories 4 and 5) from 29 patients. These 29 patients may have nodules with benign features concurrently. The range of patients' ages is between 23 and 87 years, and the average age is 52.9. Each breast has three scans at maximum to capture the entire breast region and scans with at least one breast nodule (BI-RADS 2, 3, 4, and 5) were included in the in-house ABUS dataset A. Because we limit our study to detect nodules within BI-RADS categories 3, 4, and 5 (55, 90, and 19 nodules, respectively) due to their clinical impact, breast nodules with BI-RADS category 2 is not the target of detection. Radiologist A (with four years of experience) participated in annotating breast nodules in the in-house ABUS datasets A. We used the in-house ABUS dataset A for model development and cross-validation tests.

Additionally, the in-house ABUS dataset B has been acquired from February to June 2021 as a separate test dataset, which is an alternative to an external validation dataset under our limited circumstances. The in-house ABUS dataset B consists of 168 scans from 28 female patients. It contains 126 scans with benign-appearing nodules (BI-RADS categories 2 and 3) from 21 female patients and 42 scans with suspicious nodules (BI-RADS categories 4 and 5) from 7 female patients. The six scans covering the entire breast regions per patient were included in the in-house ABUS dataset B. Hence, it may contain scans capturing the breast region without nodules, unlike the in-house ABUS dataset A. The in-house ABUS dataset B includes 23, 34, and 6 nodules in BI-RADS categories 3, 4, and 5, respectively. Radiologist B (with 27 years of experience) participated in annotating breast nodules in the in-house ABUS dataset B to secure diversity and prevent possible bias. Moreover, an auxiliary ABUS dataset was acquired in February 2023 for chest wall region segmentation to remove FPs using the U-Net model. The auxiliary ABUS dataset consists of 480 images in the transverse plane from 8 subjects, where 10 images were randomly chosen from each scan. Radiologist A participated in marking the boundary of the chest wall area.

Faster R-CNN training

In our application scenario, we utilize 2-D images in the transverse, coronal, and sagittal planes separately to build the breast nodule detector. For this, we adopt the Faster R-CNN due to its impressive speed and accuracy performance demonstrated in the object detection tasks in medicine[29,41–44]. The Faster R-CNN attempts to solve the computational bottleneck of the object detection networks based on region proposal algorithms[19,45]. Specifically, a region proposal network (RPN) and an object detection network are trained end-to-end via a simple alternating optimization by sharing convolutional features. This aspect enables nearly cost-free region proposals. Figure 2 illustrates a simplified flow diagram of the Faster R-CNN applied in breast nodule detection. According to[46], region proposal-based object detectors such as the Faster R-CNN tend to be more efficient in speed than those based on the sliding window and CNNs. In order to reduce the burden of retaining a large number of training examples while increasing the generalization capability, we utilized a ResNet-50 model[45] pre-trained using the Microsoft COCO dataset[47] as the backbone CNN architecture. With such setting, the Faster R-CNN can be effectively trained using relatively small data. The standard nine anchor boxes are utilized for the RPN. For optimization, we adopted the stochastic gradient descent[48] with a learning rate of 0.001 and momentum of 0.9. The learning rate was set to diminish by a factor of 0.1 for every three epochs. We set the number of epochs to 25 and the weight decay factor to 0.0005.

Since the in-house ABUS dataset consists of volumetric image data, the Faster R-CNN detector produces multiple 2-D detection outcomes for each slice. Hence, locating a 3-D breast nodule essentially requires additional processing, which are summarized as (1) removing FP bounding boxes in 2-D slice images and aggregating redundant 2-D bounding boxes, (2) establishing links between 2-D bounding boxes from sequential slices and generating 3-D cuboids, and (3) removing FP cuboids in chest wall regions and aggregating redundant 3-D cuboids.

2-D false positives reduction

Because the 2-D detection outcomes include bounding boxes with significantly small confidence values, a threshold τ_c is defined to remove weak detection outcomes considered FPs. Subsequently, we utilize hierarchical clustering[49] to aggregate redundant 2-D bounding boxes detected in a single slice image. As a similarity metric between two bounding boxes, the intersection over union (IoU) criterion[50] is utilized in performing the hierarchical clustering technique. As a cut-off value, we set 0.3 to decide the bounding boxes are associated. Consequently, the representative bounding box of a cluster is defined by calculating the median coordinates of associated bounding boxes. The median operation is performed per each corner of the associated bounding boxes.
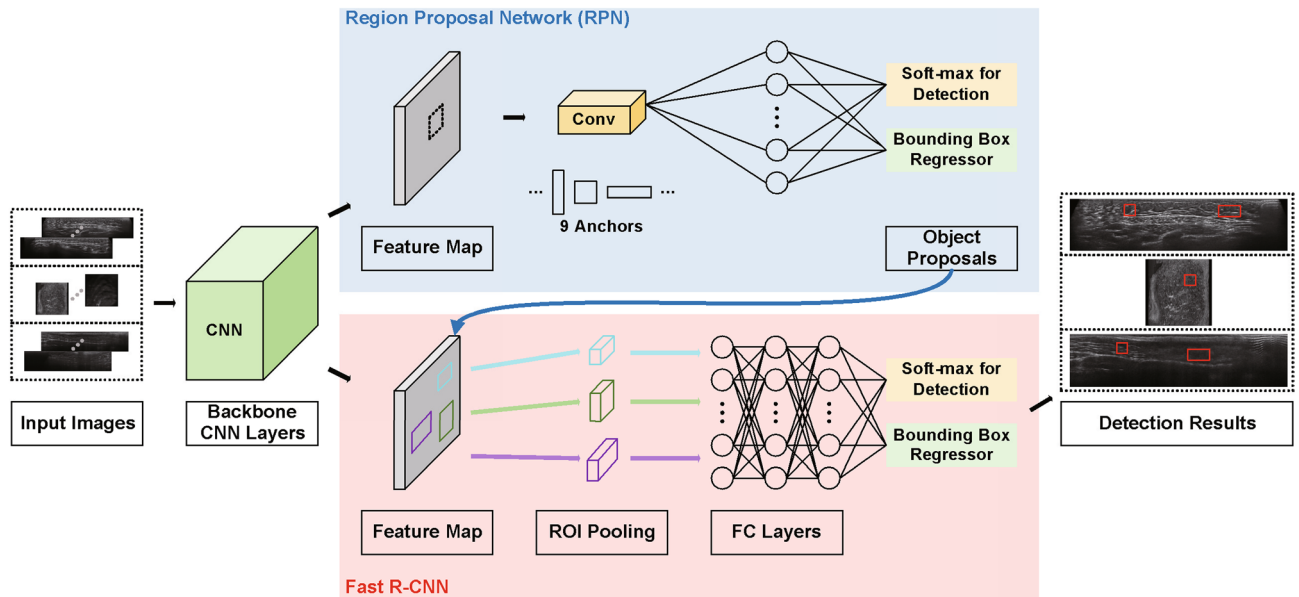
Figure 2. A simplified diagram of the Faster R-CNN.

2-D bounding boxes to 3-D cuboids conversion

The 3-D breast nodule cuboid is located by aggregating 2-D bounding box information extracted from 2-D image slices. Considering every pair of adjacent 2-D slice images, sequential 2-D bounding boxes with an IoU greater than or equal to a threshold τ_i are regarded as connected. If the number of slices with connected bounding boxes is higher than or identical to a threshold τ_s, the sequential 2-D bounding boxes are converted to a 3-D cuboid. Otherwise, they are assumed small and excluded for further processing. Because the connected 2-D bounding boxes are located irregularly throughout the sequential slice images, 3-D cuboid information is constructed by obtaining 2-D median row and column coordinates of the four corners of connected 2-D bounding boxes and the minimum and maximum slice numbers. The resulting 3-D cuboid information is described using the minimum and maximum values of row, column, and slice axes. Figure 3 illustrates a flow diagram for the 2-D bounding boxes to the 3-D cuboid conversion process.

3-D false positives reduction

Because generated 3-D breast nodule cuboids may include redundant outcomes and FPs, we perform an additional FP reduction process based on chest wall region segmentation. For this purpose, the U-Net model[15] with a residual network with 18-layer (ResNet-18)[26] pre-trained on the ImageNet dataset[51] as the encoder is utilized. The model training was performed using the auxiliary ABUS in-house dataset, where a sensitivity, specificity,
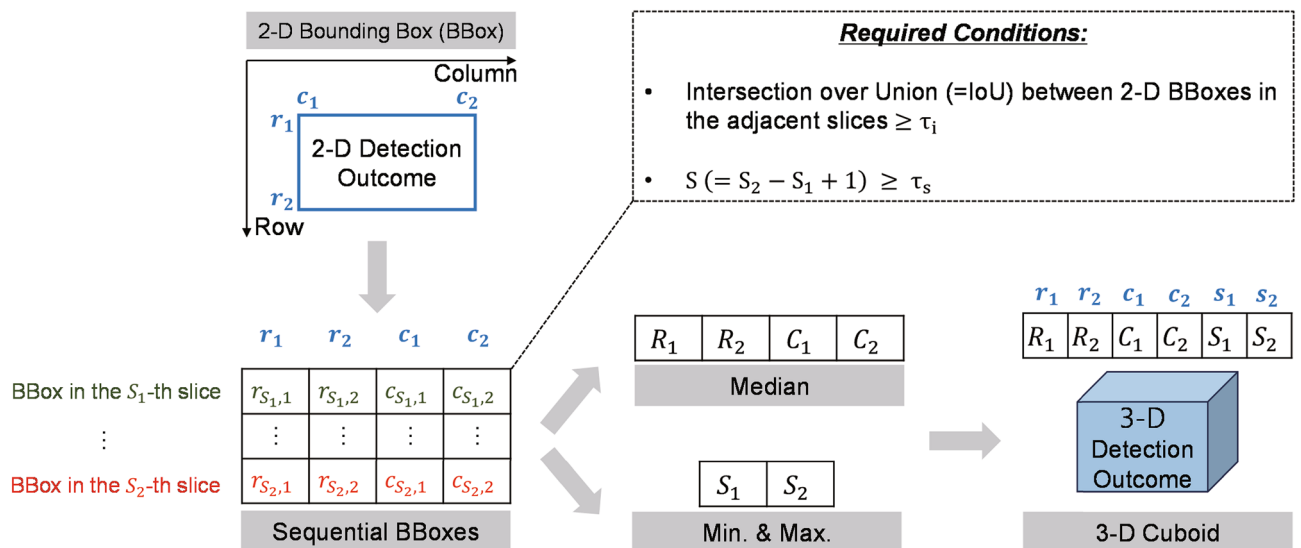

Figure 3. A flow diagram for 2-D bounding boxes to 3-D cuboids conversion process.

and accuracy of 86.77 ± 4.15%, 91.75 ± 2.12%, and 89.99 ± 2.48%, respectively were achieved from a single run of subject-wise eight-fold cross-validation tests. In order to apply the model for the in-house ABUS datasets A and B, re-training was performed using the entire auxiliary dataset to secure the highest number of training samples. Those 3-D breast nodule cuboids with center coordinates located in the segmented chest wall regions are regarded as FPs and excluded for further processing. Additionally, redundant 3-D cuboids are aggregated using a hierarchical clustering technique with setting IoU criterion in a volumetric manner as a similarity metric. The aggregation procedure is similar to those in 2-D FP reduction but in a 3-D sense. Consequently, the resulting 3-D breast nodule cuboids are defined as final 3-D breast nodules detected and used for performance assessment.

Results

Evaluation protocols

In our experiments, breast nodule detection performance is assessed for ABUS images in BI-RADS categories 3, 4, and 5. Firstly, we performed a single run of ten-fold cross-validation tests using the in-house ABUS dataset A. During the training phase, a single run of ten-fold hold-out validation was performed using the training set. The data partitioning tasks were performed based on the patient. Secondly, we performed a hold-out test using the best model determined from the cross-validation tests using the in-house ABUS dataset B. To keep as much as initial detection outcomes and remove undesirable outcomes at false positives reduction stages, we set thresholds τ_i and τ_s at 0.3 and 5 (corresponding to 1 mm) heuristically during the 2-D bounding boxes to 3-D cuboids conversion stage.

In our study, a detection outcome was considered a true positive (TP) if the center coordinate of the detected cuboid is within a ground truth (GT) cuboid according to Yap et al.[24]. There are other approaches utilizing the distance between the center coordinates of TP and GT cuboids (or bounding boxes for 2-D detection). For example, Chiang et al.[32] and Moon et al.[31] set 10 mm and 15 mm as the criteria. Hence, we also present the average and standard deviation of the distance between the center coordinates of TP and GT cuboids. For performance metrics, we employed the sensitivity (%), the number of TP nodules (#TPs), the number of GT nodules (#GTs), precision (%), the number of false positives (FPs) per pass, and the number of detected cuboids (#DETs) per pass. Performance evaluation using these metrics was performed for (1) the entire data, (2) each BI-RADS category (BI-RADS categories 3, 4, and 5, separately), (3) a group of BI-RADS categories (suspicious nodules in BI-RADS categories 4 and 5), and (4) data without targetted breast nodules for detection (controlled group). A single run of cross-validation tests with a single run of hold-out validation using only the training set was performed for the in-house ABUS dataset A. The in-house ABUS dataset B was utilized as a separate test dataset to deal with the lack of data for external validation, where a single run of hold-out test was performed for the in-house ABUS dataset B. For the results from the cross-validation tests, a paired t-test[52,53] was performed to demonstrate whether the improvement via 3-D FPs reduction is statistically significant (significance level = 0.01), and the differences among three plane images are statistically significant (significance level = 0.05). The statistical significance test was not performed for the hold-out test using the in-house ABUS dataset B because we obtained a single set of results from the test.

Confidence threshold value selection

To decide the confidence threshold value, we set a range between 0 and 0.8 with an increment of 0.1 and evaluated the detection performance. Figure 4 shows average values of the sensitivity and the number of FPs per pass measured at the validation phase during the cross-validation tests. Considering the trade-off between the sensitivity and the number of FPs, we selected the confidence threshold value τ_c, when the number of FPs is less

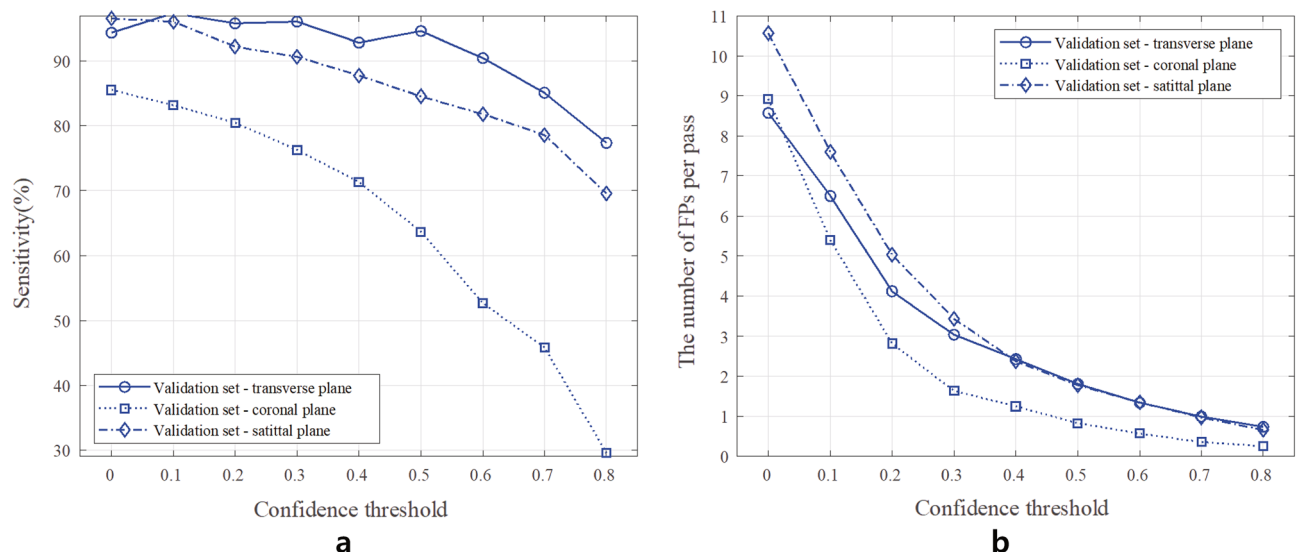**a**   **b**

Figure 4. The sensitivity and number of FPs per pass with respect to the confidence threshold from experiments using the in-house ABUS datasets A and B. (**a**) sensitivity (%). (**b**) the number of FPs per pass.

than 2. As a result, the confidence threshold τ_c was chosen at 0.5 for transverse and sagittal plane images. Because the coronal plane images showed much lower sensitivity than those from other plane images at $\tau_c = 0.5$, the confidence threshold for the coronal plane images was selected using the sensitivity values of the transverse and sagittal plane images at $\tau_c = 0.5$. Consequently, a sensitivity of around 85%, which corresponds to the minimum value among those of the other plane images, was utilized to choose τ_c for the coronal plane images. These result in obtaining $\tau_c = 0.5$ for the transverse and sagittal plane images and $\tau_c = 0$ for the coronal plane images, respectively. Hereafter, the breast performance evaluation and analysis regarding performance enhancement via chest wall region segmentation were based on the selected operating point of the confidence values.

Comprehensive performance evaluation

The sensitivity values and the number of TP and GT nodules measured from the cross-validation tests using the in-house ABUS datasets A detecting breast nodules in BI-RADS 3, 4, and 5 are presented in Table 1. The same sensitivity values were obtained before and after applying the 3-D FPs reduction for the entire experiment. Table 2 shows the number of FPs per pass and the precision values measured using the in-house datasets A. Due to the trade-off between the sensitivity and the precision, low precision values were obtained at the selected operating point. Hence, we utilize the number of FPs per pass directly to discuss how precisely our model detects breast nodules, and also present the number of detected cuboids. For all cases, the performance improvement of the number of FPs per pass by 3-D FPs reduction was statistically significant (significance level = 0.01). From the cross-validation tests using the in-house ABUS dataset A, the transverse plane images showed an average sensitivity of 90.98% at 11.6 average FPs per pass after applying 3-D FPs reduction. There exist 15.2 ground-truth nodules and 12.2, 15.4, and 9.6 detected cuboids per plane in the test sets on average. The coronal (84.41% sensitivity at 14.7 FPs per pass) and sagittal (83.38% at 8.9 FPs per pass) plane images showed lower sensitivity than the transverse plane images. Among the three plane images, the transverse plane images showed a significantly better sensitivity performance than the sagittal plane images (significance level = 0.05). The other sensitivity comparisons between planes were not statistically significant. The sagittal plane images showed a significantly

Nodule category	ABUS dataset	Sensitivity (%)			#TPs / #GTs		
		Transverse	Coronal	Sagittal	Transverse	Coronal	Sagittal
Total	A	90.98 ± 7.29	84.41 ± 13.20	83.38 ± 10.38	13.9 / 15.2	12.8 / 15.2	12.7 / 15.2
	B	93.65	87.30	87.30	59 / 63	55 / 63	55 / 63
BI-RADS 3	A	85.24 ± 21.44	80.79 ± 19.16	79.64 ± 16.91	4.1 / 4.7	3.8 / 4.7	3.8 / 4.7
	B	91.30	78.26	69.57	21 / 23	18 / 23	16 / 23
BI-RADS 4	A	88.73 ± 20.46	85.09 ± 15.98	82.72 ± 14.85	7.6 / 8.3	6.8 / 8.3	6.8 / 8.3
	B	94.12	91.18	97.06	32 / 34	31 / 34	33 / 34
BI-RADS 5	A	100.00 ± 0.00	100.00 ± 0.00	94.44 ± 16.67	2.2 / 2.2	2.2 / 2.2	2.1 / 2.2
	B	100.00	100.00	100.00	6 / 6	6 / 6	6 / 6
Suspicious	A	90.91 ± 15.41	87.55 ± 14.19	85.13 ± 11.02	9.8 / 10.5	9.0 / 10.5	8.9 / 10.5
	B	95.00	92.50	97.50	38 / 40	37 / 40	39 / 40

Table 1. Sensitivity (%), the number of true positive nodules (#TPs), the number of ground-truth nodules (# GTs) measured using the in-house ABUS datasets A and B. Performance values are reported for a set of nodule groups, such as total nodules (BI-RADS categories 3, 4, and 5), each BI-RADS category, and suspicious nodules (BI-RADS categories 4 and 5). Mean (sensitivity, #TPs, and #GTs) and standard deviation (sensitivity) values are reported for the in-house ABUS dataset A, and a single value is reported for the in-house ABUS dataset B.

Nodule category	3-D FPs reduction	ABUS dataset	#FPs per pass (#DETs per pass)			Precision (%)		
			Transverse	Coronal	Sagittal	Transverse	Coronal	Sagittal
Total	No	A	14.1 ± 5.6 (14.7 ± 5.7)	24.1 ± 12.2 (24.7 ± 12.2)	11.9 ± 3.7 (12.6 ± 3.8)	4.69 ± 2.56	3.30 ± 2.08	5.56 ± 2.96
		B	10.8 (11.2)	13.9 (14.4)	16.6 (17.0)	4.13	3.77	3.25
	Yes	A	$11.6 \pm 5.6^*$ (12.2 ± 5.7)	$14.7 \pm 6.3^*$ (15.4 ± 6.3)	$8.9 \pm 4.3^*$ (9.6 ± 4.5)	5.81 ± 3.09	4.88 ± 2.63	7.49 ± 3.47
		B	8.6 (9.0)	9.9 (10.4)	16.0 (16.4)	5.18	5.16	3.38

Table 2. The number of false positives (#FPs) per pass, the number of detected cuboids (#DETs) per pass, and the precision (%) measured using the in-house ABUS datasets A and B for detection of total nodules (BI-RADS categories 3, 4, and 5). Mean and standard deviation values are reported for the in-house ABUS dataset A, and a single value is reported for the in-house ABUS dataset B. The symbol * denotes that the performance improvement via 3-D FPs reduction is statistically significant ($p < 0.001$).

smaller number of FPs per pass than the coronal plane images (significance level = 0.05), while there was no statistically significant difference for the remaining comparisons between planes.

The hold-out test results using the in-house ABUS dataset B for detecting breast nodules in BI-RADS categories 3, 4, and 5 are presented in Tables 1 and 2. The transverse plane images showed the best results (93.65% sensitivity at 8.6 FPs per pass). The coronal plane images (87.30% sensitivity at 9.9 FPs per pass) showed smaller FPs per pass than the sagittal plane images (87.30% sensitivity at 16.0 FPs per pass). There are 63 GT nodules in total, and 9.0, 10.4, and 16.4 detected cuboids for the transverse, coronal, and sagittal planes. The 3-D FPs reduction produced decrease of around 2.2, 4.0, and 0.6 FPs per pass for the three planes. In addition to the performance assessment, the average distance between the center coordinates of TP and GT cuboids for the in-house ABUS datasets A and B is presented in Table 3. The average distance ranges from 3.50 mm to 4.55 mm for the in-house ABUS dataset A and 3.25 mm to 4.37 mm for the in-house ABUS dataset B. There values are less than the criteria reported in the literature (e.g., 10 mm[32] and 15 mm[31]). Comprehensively less than 4.24% of the entire slices in each volume contain TP or FP detection outcomes for the three views on average (in-house ABUS dataset A: 2.39–2.99%, and in-house ABUS dataset B: 1.42–4.24%). Hence, given the sensitivity performance shown in Table 1, the radiologists need to read less than 5% of the entire images in each volume for interpretation. Applying the proposed CADe system may result in assistance via obtaining candidate nodule regions in an automatic manner and conserving reading time drastically.

Performance evaluation per BI-RADS category

The sensitivity and the number of TP and GT nodules per each BI-RADS category measured using the in-house datasets A are presented in Table 1. Table 4 shows the number of FPs and DETs per pass and the precision

ABUS Dataset	Distance between Centers of TP and GT Cuboids (mm)		
	Transverse	Coronal	Sagittal
A	3.50 ± 2.45	4.03 ± 3.15	4.55 ± 3.21
B	3.25 ± 2.40	4.20 ± 3.72	4.37 ± 3.07

Table 3. The distance between the center coordinates of true positive (TP) and ground truth (GT) cuboids in millimeters for BI-RADS categories 3, 4, and 5 nodule detection. Mean and standard deviation values are reported.

Nodule Category	3-D FPs Reduction	ABUS Dataset	#FPs per Pass (#DETs per Pass)			Precision (%)		
			Transverse	Coronal	Sagittal	Transverse	Coronal	Sagittal
BI-RADS 3	No	A	15.4 ± 7.4 (16.7 ± 7.1)	23.9 ± 13.3 (25.1 ± 13.5)	12.4 ± 4.9 (13.7 ± 5.2)	10.64 ± 10.25	6.01 ± 4.23	9.91 ± 4.83
		B	11.1 (12.2)	16.6 (17.5)	19.2 (20.1)	9.13	5.37	4.59
	Yes	A	$13.3 \pm 7.0^*$ (14.6 ± 6.8)	$14.1 \pm 8.5^*$ (15.3 ± 8.6)	$9.3 \pm 5.2^*$ (10.6 ± 5.4)	12.35 ± 11.20	17.68 ± 29.54	13.52 ± 6.51
		B	15.6 (16.8)	18.1 (19.1)	23.9 (25.1)	7.30	5.46	4.88
BI-RADS 4	No	A	13.7 ± 10.1 (15.1 ± 10.7)	21.4 ± 10.1 (23.2 ± 9.8)	9.2 ± 4.4 (10.8 ± 4.7)	16.15 ± 21.18	8.96 ± 5.30	16.40 ± 8.85
		B	12.0 (13.5)	16.7 (18.5)	20.2 (22.2)	11.08	9.91	8.86
	Yes	A	$12.0 \pm 10.4^*$ (13.4 ± 11.0)	$12.4 \pm 5.7^*$ (14.1 ± 5.5)	$6.9 \pm 4.1^*$ (8.4 ± 4.5)	20.37 ± 28.57	14.30 ± 8.13	20.56 ± 8.10
		B	10.4 (11.9)	13.3 (15.1)	19.7 (21.6)	12.61	12.14	9.09
BI-RADS 5	No	A	14.1 ± 10.5 (15.8 ± 10.5)	27.1 ± 17.3 (29.3 ± 18.1)	11.0 ± 9.4 (13.4 ± 10.1)	18.63 ± 18.67	8.57 ± 4.52	22.44 ± 13.90
		B	4.5 (5.5)	11.0 (13.0)	11.5 (12.7)	18.18	15.38	9.21
	Yes	A	$13.2 \pm 9.7^*$ (15.0 ± 9.7)	$19.3 \pm 13.2^*$ (21.5 ± 13.9)	$9.8 \pm 8.3^*$ (12.1 ± 9.0)	19.11 ± 18.53	12.42 ± 6.82	24.32 ± 14.70
		B	3.8 (4.8)	7.3 (9.3)	11.0 (12.2)	20.69	21.43	9.59

Table 4. The number of false positives (#FPs) per pass, the number of detected cuboids (#DETs) per pass, and the precision (%) measured using the in-house ABUS datasets A and B per volumes in each BI-RADS category. Mean and standard deviation values are reported for the in-house ABUS dataset A, and a single value is reported for the in-house ABUS dataset B. The symbol * denotes that the performance improvement via 3-D FPs reduction is statistically significant ($p < 0.001$).

measured from the cross-validation tests using the in-house ABUS dataset A per each BI-RADS category. As the results for the total BI-RADS categories, no sensitivity performance degradation was observed for the 3-D FPs reduction, and the performance improvement by the 3-D FPs reduction for all cases was statistically significant (significance level = 0.01). As the BI-RADS category increases, the sensitivity performance increases for all planes. For the cross-validation tests using the in-house ABUS datasets A, we obtained average sensitivities of 85.24% at 13.3 FPs per pass (transverse, 14.6 detected cuboids), 80.79% at 14.1 FPs per pass (coronal, 15.3 detected cuboids), and 79.64% at 9.3 FPs per pass (sagittal, 10.6 detected cuboids) for volumes with BI-RADS category 3. For BI-RADS category 4, sensitivities of 88.73% at 12.0 FPs per pass (transverse, 13.4 detected cuboids), 85.09% at 12.4 FPs per pass (coronal, 14.1 detected cuboids), and 82.72% at 6.9 FPs per pass (sagittal, 8.4 detected cuboids) were observed on average. For volumes with nodules in BI-RADS category 5, all GT nodules were successfully detected for the transverse and coronal planes with 13.2 and 19.3 FPs per pass among 15.0 and 21.5 detected cuboids, respectively. At the same time, a sensitivity of 94.44% at 9.8 FPs per pass was achieved for the sagittal plane images (12.1 detected cuboids). For each category, the sagittal plane images outperformed the coronal plane images in terms of the number of FPs per pass with statistical significance (significance level = 0.05). There was no other statistically significant difference between planes.

From the hold-out test using the in-house ABUS dataset B, sensitivities of 91.30% at 15.6 FPs per pass, 78.26% at 18.1 FPs per pass, and 69.57% at 23.9 FPs per pass were obtained for the images of the transverse, coronal, and sagittal planes in BI-RADS category 3 (Tables 1 and 4). There were 16.8, 19.1, and 25.1 detected cuboids for the transverse, coronal, and sagittal planes, and 23 GT nodules for BI-RADS category 3. For scans with nodules in BI-RADS category 4, we achieved sensitivities of 94.12% at 10.4 FPs per pass (transverse, 11.9 detected cuboids), 91.18% at 13.3 FPs per pass (coronal, 15.1 detected cuboids), and 97.06% at 19.7 FPs per pass (sagittal, 21.6 detected cuboids). All GT nodules within BI-RADS category 5 (6 GT nodules) were successfully detected with 3.8, 7.3, and 11.0 FPs per pass among 4.8, 9.3, and 12.2 detected cuboids for the transverse, coronal, and sagittal planes. As in the cross-validation tests using the in-house ABUS dataset A, the sensitivity value rises as the BI-RADS category increases.

Performance evaluation for volumes with suspicious nodules

The sensitivity and the number of TP and GT nodules measured from the cross-validation tests using the in-house ABUS dataset A for detecting suspicious nodules (BI-RADS categories 4 and 5) are reported in Table 1. Table 5 presents the number of FPs per pass, the number of detected cuboids per pass, and the precision measured from the cross-validation tests. The 3-D FPs reduction produced a statistically significant reduction of FPs (significance level = 0.01) without degrading the sensitivity for all cases. Sensitivities of 90.91% at 12.1 FPs per pass (transverse, 13.6 detected cuboids), 87.55% at 13.8 FPs per pass (coronal, 15.7 detected cuboids), and 85.13% at 7.1 FPs per pass (sagittal, 8.9 detected cuboids) were observed, while there was no statistically significant difference between the sensitivities between planes. In terms of FPs per pass, the sagittal plane images significantly outperformed the coronal plane images (significance level = 0.05). There was no other statistically significant difference in FPs per pass between planes.

The sensitivity, the number of FP and GT nodules measured from the hold-out test using the in-house ABUS dataset B for the suspicious nodules are presented in Table 1. The number of FPs per pass, the number of detected cuboids per pass, and the precision corresponding to the sensitivity value are provided in Table 5. As in the cross-validation tests, the 3-D FPs reduction retained the original sensitivity performance. For 40 GT nodules with suspicious nodules (BI-RADS categories 4 and 5), sensitivities of 95.00% at 9.3 FPs per pass (transverse, 10.7 detected cuboids), 92.50% at 12.3 FPs per pass (coronal, 14.1 detected cuboids), and 97.50% at 18.2 FPs per pass (sagittal, 20.1 detected cuboids). Compared to the results for scans with benign-appearing nodules in BI-RADS category 3, detecting nodules with suspicious features is regarded to be relatively easy due to their distinct appearances.

Nodule Category	3-D FPs Reduction	ABUS Dataset	#FPs per Pass (#DETs per Pass)			Precision (%)		
			Transverse	Coronal	Sagittal	Transverse	Coronal	Sagittal
Suspicious	No	A	13.6 ± 10.0 (15.1 ± 10.4)	22.5 ± 11.3 (24.4 ± 11.2)	9.1 ± 4.5 (10.8 ± 4.9)	16.79 ± 19.75	9.01 ± 4.94	18.48 ± 8.86
		B	10.8 (12.2)	15.7 (17.6)	18.8 (20.6)	11.62	10.58	9.89
	Yes	A	12.1 ± 10.0* (13.6 ± 10.4)	13.8 ± 6.9* (15.7 ± 6.8)	7.1 ± 4.1* (8.9 ± 4.5)	19.56 ± 23.17	13.99 ± 7.62	22.39 ± 8.34
		B	9.3 (10.7)	12.3 (14.1)	18.2 (20.1)	13.21	13.16	9.14

Table 5. The number of false positives (#FPs) per pass, the number of detected cuboids (#DETs) per pass, and the precision (%) measured using the in-house ABUS datasets A and B for volumes with suspicious nodules (BI-RADS categories 4 and 5) Mean and standard deviation values are reported for the in-house ABUS dataset A, and a single value is reported for the in-house ABUS dataset B. The symbol * denotes that the performance improvement via 3-D FPs reduction is statistically significant ($p < 0.001$).

Performance evaluation for controlled group

Our study aims to detect breast nodules within BI-RADS categories 3, 4, and 5 due to their clinical impact. Because the in-house ABUS datasets A and B contain nodules in BI-RADS category 2, and the in-house ABUS dataset B includes scans without nodules, subsets of them are utilized as a controlled group. In these subsets, all detection outcomes are considered FPs. Table 6 presents the number of FPs per pass measured using the subsets. The performance improvement via 3-D FPs reduction was statistically significant (significance level = 0.01) for all plane images of the in-house ABUS dataset A. For 17 scans in the in-house ABUS dataset A with breast nodules in BI-RADS category 2, 5.8, 9.6, and 6.6 FPs per pass were obtained for the transverse, coronal, and sagittal planes by 3-D FPs reduction. We achieved 8.1 (transverse), 9.8 (coronal), and 15.5 (sagittal) FPs per pass for 36 scans in the in-house ABUS dataset B with breast nodules in BI-RADS category 2. For 80 scans without nodules in the in-house ABUS dataset B, 7.3, 7.5, and 13.6 FPs per pass were observed for the transverse, coronal, and sagittal planes. These values are slightly lower than those measured for total scans, as reported in Table 2.

Detection outcomes visualization

The TP results on the in-house ABUS dataset A and B are illustrated in Figs. 5 and 6, respectively. The nodules with relatively large sizes are easily distinguishable from the background tissue (see BI-RADS categories 4 and 5 cases in Figs. 5 and 6). Even the relatively small nodules in BI-RADS category 3 show clear separation from the background tissue for TP results. Figure 7 illustrates the FP results on the in-house ABUS datasets. The FP

3-D FPs reduction	ABUS dataset	The number of false positives (FPs) per pass		
		Transverse	Coronal	Sagittal
No	A1	9.0 ± 5.1	15.4 ± 3.1	11.8 ± 7.6
	B1	9.6	13.1	16.0
	B2	10.5	12.1	14.4
Yes	A1	5.8 ± 3.6*	9.6 ± 3.3*	6.6 ± 4.4*
	B1	8.1	9.8	15.5
	B2	7.3	7.5	13.6

Table 6. The number of false positives (#FPs) per pass measured using partial volumes of the in-house ABUS datasets A and B. A1 (17 scans) and B1 (36 scans) are subsets of the in-house ABUS datasets A and B, respectively, where the scans contain nodules solely in BI-RADS category 2. B2 (80 scans) denotes a subset of the in-house ABUS dataset B, where the scans have no nodule. The symbol * denotes that the performance improvement via 3-D FPs reduction is statistically significant ($p < 0.001$).
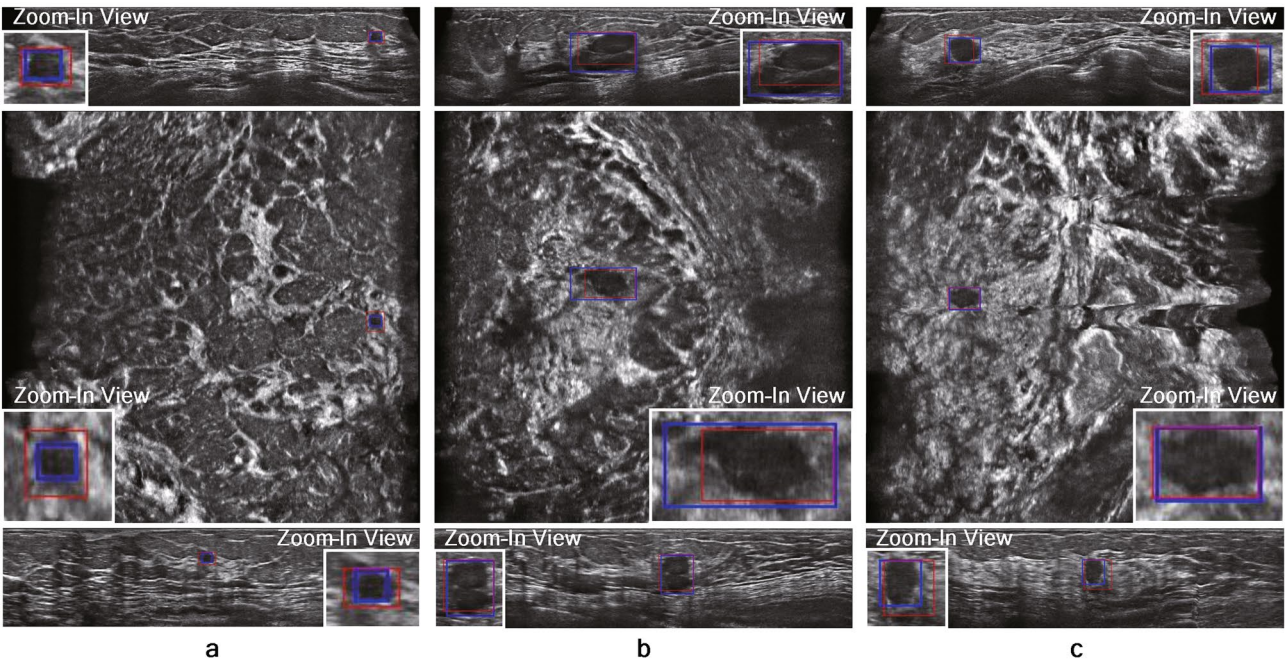

Figure 5. TP results on the in-house ABUS dataset A. (**a**) BI-RADS category 3 case. (**b**) BI-RADS category 4 case. (**c**) BI-RADS category 5 case. Images are arranged in the order of transverse, coronal, and sagittal views in a downward direction. Blue boxes with a thick solid line show the ground truth bounding box, and red boxes with a thin solid line show the TP detection.
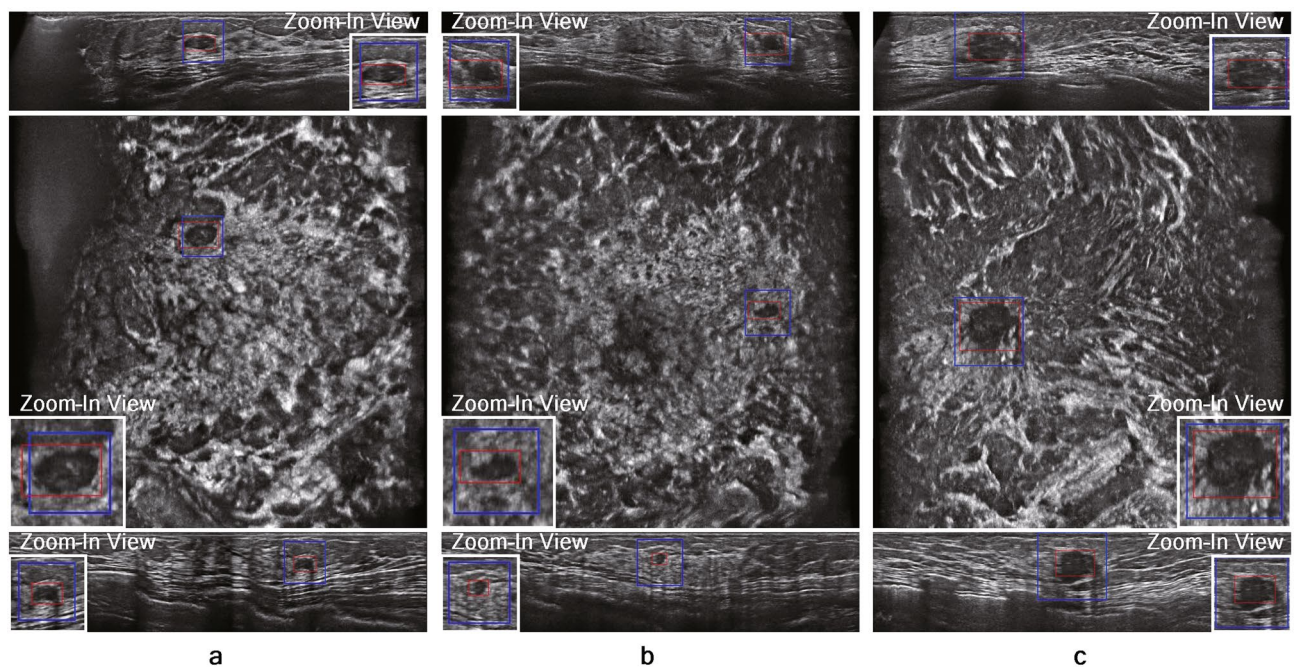
Figure 6. TP results on the in-house ABUS dataset B. (**a**) BI-RADS category 3 case. (**b**) BI-RADS category 4 case. (**c**) BI-RADS category 5 case. Images are arranged in the order of transverse, coronal, and sagittal views in a downward direction. Blue boxes with a thick solid line show the ground truth bounding box, and red boxes with a thin solid line show the TP detection.
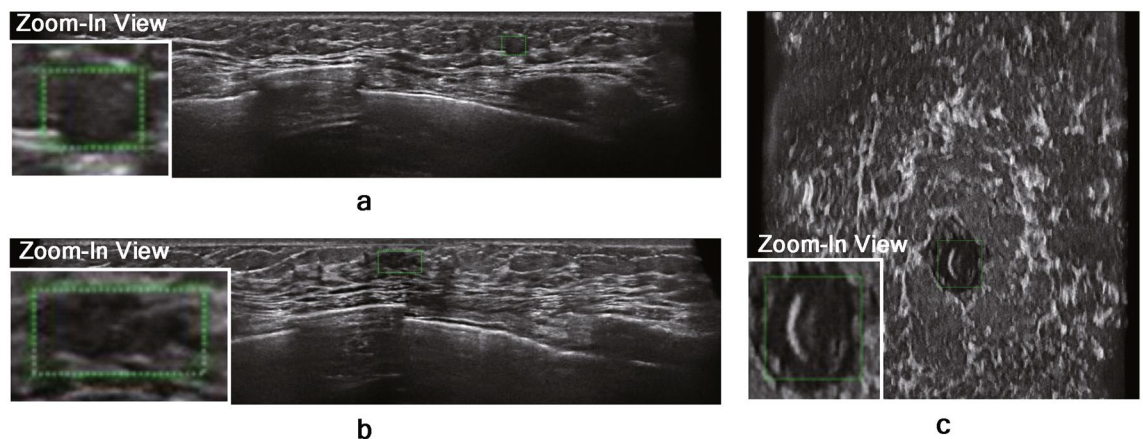

Figure 7. FP results on the in-house ABUS dataset A and B. (**a**) fat lobule (transverse view). (**b**) acoustic shadowing (transverse view). (**c**) nipple (coronal view). Green boxes with a dotted line show the FP detection.

results include the detection of fat lobules, acoustic shadowing, and nipple, where the fat lobules and acoustic shadowing are the majority of FPs.

Discussion

In this study, we proposed a breast nodule detection system on ABUS volumes based on the Faster R-CNN and U-Net models. The breast nodules were detected in the form of a 3-D cuboid utilizing each of the transverse, coronal, and sagittal plane images separately. The hierarchical clustering is utilized to reduce 2-D FPs. Subsequently, 3-D cuboids are generated from sequential 2-D bounding boxes based on a set of operations. Consequently, the 3-D FPs in the chest wall regions are removed from the detection outcomes. From the cross-validation tests using the in-house ABUS dataset A containing breast nodules with BI-RADS categories 3, 4, and 5, we achieved average sensitivities of 90.98%, 84.41%, and 83.38% for transverse, coronal, and sagittal plane images, respectively. At such sensitivity values, 11.6, 14.7, and 8.9 FPs per pass were achieved on average among 12.2, 15.4, and 9.6 detected cuboids. For the cross-validation tests, the improvement of the number of FPs per pass via 3-D FPs reduction was statistically significant without deteriorating the sensitivity. Additionally, the transverse plane images outperformed the sagittal plane images with statistical significance in terms of sensitivity. In view of the number of FPs per pass, the sagittal plane images showed better performance than the coronal plane images

with statistical significance. The same statistical significance test results were obtained for data in each BI-RADS category and with suspicious nodules. For the hold-out test using the in-house ABUS dataset B, we achieved sensitivities of 94.65%, 87.30%, and 87.30% for the transverse, coronal, and sagittal plane images. Concurrently, 8.6, 9.9, and 16.0 FPs per pass were obtained among 9.0, 10.4, and 16.4 detected cuboids, respectively, at such sensitivity values. The average distances between TP and GT cuboids were within the ranges [3.50, 4.55] and [3.25, 4.37] millimeters for the in-house ABUS datasets A and B, where the values are smaller than the criteria reported in the literature[31,32]. For both datasets, detecting breast nodules with a higher BI-RADS category was relatively easy due to their distinct appearance. Finally, the performances result in less than 5% of the entire slices in a volume containing detection outcomes regardless of TPs or FPs. Hence, the proposed 3-D breast nodule detection system can drastically assist in saving time and effort for reading by radiologists.

Comparisons with the state-of-the-art 3-D breast nodule detection methods in the literature in terms of the sensitivity and the number of FPs per pass are presented in Table 7. Methods based on 3-D CNN (Chiang et al.[32] and Moon et al.[31]) show a certain level of good sensitivity performance (over 95.00%), while a relatively high FPs per pass (21.6) was reported by Moon et al.[31]. Compared with the study by Chiang et al.[32] (2.3 FPs per pass), our method for data with BI-RADS categories 3, 4, and 5 shows lower sensitivity and higher FPs per pass. However, they included a much larger number of malignant tumors than benign tumors, where the detection complexity is considered relatively low. Similarly, we obtained lower sensitivity and the number of FPs per pass than the study by Zhou et al.[38] based on a modified Faster R-CNN model, where the average tumor volume (2 cm^3) is considered high, and the number of patients is relatively small. Compared with the convolutional BiLSTM proposed by Malekmohammadi et al.[37], we obtained lower sensitivity with smaller FPs per pass for total data. However, they utilized a small dataset containing a much larger number of malignant nodules than benign nodules. Zhang et al.[35] reported internal and external validation results by applying the YOLOv5 model on the biggest dataset consisting of a much larger number of benign nodules than malignant nodules. Overall, our method obtained better sensitivity performances than the study by Zhang et al.[35], but with higher FPs per pass. In view of the dataset composition, our study is considered to handle a 3-D breast nodule detection problem with relatively higher complexity than other studies except for the study by Zhang et al.[29].

One of the limitations of our study is that our system tends to generate a substantial number of FPs even within the controlled group (5.8–15.5 FPs) because the system is tailored towards generating high amounts of detections. We note that this tendency may not pose a significant issue at the system deployment stage since the system is intended to be utilized by radiologists to reduce the interpretation time primarily, and potential adjustments of the operating threshold can be made for specific application scenarios. Another limitation of our

Method	#Patients	Ground truth	Validation	Nodule	SEN (%)	#FPs per Pass	Note
Chiang et al.[32]	187	Pathology-proven	Internal	Benign+Malignant	80.00	0.6	60 benign and 240 malignant tumors
					95.00	2.3	
Moon et al.[31]	246	Pathology-proven	Internal	Benign+Malignant	84.80	3.7	79 benign and 254 malignant tumors
					98.10	21.6	
Zhang et al.[35]	741	Pathology-proven	Internal	BI-RADS 2,3,4,5	78.10	–	2896 benign and 218 malignant lesions
				BI-RADS 2	74.70	4.0	
				BI-RADS 3	79.70		
				BI-RADS 4,5	96.50		
			External	BI-RADS 2,3,4,5	71.20	–	
				BI-RADS 2	88.40	4.0	
				BI-RADS 3	74.70		
				BI-RADS 4,5	95.80		
Malekmohammadi et al.[37]	43	Pathology-proven	Internal	Benign+Malignant	60.00	0.3	13 benign and 42 malignant mass
					100.00	16.0	
Zhou et al.[38]	75	Pathology-proven	Internal	Benign+Malignant	90.12	5.4	Average tumor volume = 2cm^3
					95.06	0.6	
Proposed	131	BI-RADS categorization	Internal	BI-RADS 3,4,5	90.98	11.6	78, 124, and 25 nodules in BI-RADS categories 3, 4, and 5
				BI-RADS 3	85.24	13.3	
				BI-RADS 4	88.73	12.0	
				BI-RADS 5	100.00	13.2	
				BI-RADS 4,5	90.91	12.1	
			External (alternative)	BI-RADS 3,4,5	93.65	10.8	
				BI-RADS 3	91.30	15.6	
				BI-RADS 4	94.12	10.4	
				BI-RADS 5	100.00	3.8	
				BI-RADS 4,5	95.00	9.3	

Table 7. Performance comparisons with state-of-the-art methodologies in the literature in terms of the sensitivity (SEN, %) and the number of false positives (#FPs per pass).

system is that there is no process to reduce FPs regarding fat lobule and acoustic shadowing. We believe filtering out most of these regions does not engage a complex methodology. Additionally, we substitute external validation using experiments on a dataset acquired from the same institute but in different acquisition periods due to a lack of data availability. Moreover, the auxiliary ABUS dataset is regarded as relatively small (eight subjects and 480 images), where retaining more data for segmenting the chest wall regions is selected as one of our immediate future works. Our future works include handling the FPs, such as fat lobule and acoustic shadowing, modifying the Faster R-CNN model to be more suitable for breast nodule detection, and performing external validation.

In this study, we proposed a 3-D breast nodule detection system using a simple 2-D Faster R-CNN model. A threshold operation on confidence values and aggregation of redundant 2-D bounding boxes were performed to reduce 2-D FPs. Subsequently, to build 3-D breast nodule information using 2-D detection outcomes, we presented 2-D bounding boxes to a 3-D cuboid conversion module. Additionally, 3-D cuboids in the chest wall region were removed based on a U-Net model. Similar to the 2-D case, redundant 3-D cuboids were aggregated to reduce 3-D FPs. Our experiments using the in-house datasets A and B show encouraging results in terms of the sensitivity and the number of FPs.

Data availability
The datasets generated during and/or analyzed during the current study are not publicly available due to privacy or ethical restrictions but are available from the first author or the corresponding author on reasonable request.

Code availability
The code for preprocessing and learning contains intellectual property and cannot be released publicly. The preprocessing module and Faster R-CNN algorithm were implemented based on https://albumentations.ai/docs/examples/pytorch_classification/ and https://pytorch.org/vision/stable/_modules/models/detection/faster_rcnn.html, respectively.

References
1. Leong, L. C. *et al.* Supplementary breast ultrasound screening in asian women with negative but dense mammograms-a pilot study. *Ann. Acad. Med. Singapore* **41**, 432 (2012).
2. Rella, R. *et al.* Automated breast ultrasonography (ABUS) in the screening and diagnostic setting: Indications and practical use. *Acad. Radiol.* **25**, 1457–1470 (2018).
3. Chang, R.-F. *et al.* Whole breast computer-aided screening using free-hand ultrasound. *Int. Congress Ser.* **1281**, 1075–1080 (2005).
4. Ikedo, Y. *et al.* Development of a fully automatic scheme for detection of masses in whole breast ultrasound images. *Med. Phys.* **34**, 4378–4388 (2007).
5. Chang, R.-F. *et al.* Rapid image stitching and computer-aided detection for multipass automated breast ultrasound. *Med. Phys.* **37**, 2063–2073 (2010).
6. Moon, W. K. *et al.* Computer-aided tumor detection based on multi-scale blob detection algorithm in automated breast ultrasound images. *IEEE Trans. Med. Imaging* **32**, 1191–1200 (2013).
7. Vourtsis, A. & Kachulis, A. The performance of 3D ABUS versus HHUS in the visualisation and BI-RADS characterisation of breast lesions in a large cohort of 1,886 women. *Eur. Radiol.* **28**, 592–601 (2018).
8. Zanotel, M. *et al.* Automated breast ultrasound: Basic principles and emerging clinical applications. *La Radiologia Medica* **123**, 1–12 (2018).
9. Brunetti, N. *et al.* Comparison between execution and reading time of 3D ABUS versus HHUS. *La Radiologia Medica* **125**, 1243–1248 (2020).
10. Jia, M. *et al.* Diagnostic performance of automated breast ultrasound and handheld ultrasound in women with dense breasts. *Breast Cancer Res. Treat.* **181**, 589–597 (2020).
11. Lin, X. *et al.* The diagnostic performance of automated versus handheld breast ultrasound and mammography in symptomatic outpatient women: A multicenter, cross-sectional study in china. *Eur. Radiol.* **31**, 947–957 (2021).
12. Boca, I., Ciurea, A. I., Ciortea, C. A. & Dudea, S. M. Pros and cons for automated breast ultrasound (abus): A narrative review. *J. Personal. Med.* **11**, 703 (2021).
13. Yap, M. H. *et al.* Automated breast ultrasound lesions detection using convolutional neural networks. *IEEE J. Biomed. Health Inform.* **22**, 1218–1226 (2018).
14. LeCun, Y., Bottou, L., Bengio, Y. & Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **86**, 2278–2324 (1998).
15. Ronneberger, O., Fischer, P. & Brox, T. U-Net: convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention, MICCAI 2015*, 234–241 (Springer, 2015).
16. Long, J., Shelhamer, E. & Darrell, T. Fully convolutional networks for semantic segmentation. *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. CVPR* **2015**, 3431–3440 (2015).
17. Krizhevsky, A., Sutskever, I. & Hinton, G. E. Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **25** (2012).
18. Cao, Z., Duan, L., Yang, G., Yue, T. & Chen, Q. An experimental study on breast lesion detection and classification from ultrasound images using deep learning architectures. *BMC Med. Imaging* **19**, 1–9 (2019).
19. Girshick, R. Fast R-CNN. *Proc. IEEE Int. Conf. Comput. Vis. ICCV* **2015**, 1440–1448 (2015).
20. Ren, S., He, K., Girshick, R. & Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks. *Adv. Neural Inf. Process. Syst.* **28**, 91–99 (2015).
21. Redmon, J., Divvala, S., Girshick, R. & Farhadi, A. You only look once: Unified, real-time object detection. *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. CVPR* **2016**, 779–788 (2016).
22. Redmon, J. & Farhadi, A. YOLOv3: an incremental improvement. *arXiv preprint* arXiv:1804.02767 (2018).
23. Liu, W. *et al.* SSD: Single shot multibox detector. In *Proceedings of the 14th European Conference on Computer Vision, ECCV 2016*, 21–37 (Springer, 2016).
24. Yap, M. H. *et al.* Breast ultrasound region of interest detection and lesion localisation. *Artif. Intell. Med.* **107**, 101880 (2020).
25. Li, Y., Gu, H., Wang, H., Qin, P. & Wang, J. BUSnet: A deep learning model of breast tumor lesion detection for ultrasound images. *Front. Oncol.* **12**, 848271 (2022).

26. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. CVPR* **2016**, 770–778 (2016).
27. Lo, C.-M. *et al.* Multi-dimensional tumor detection in automated whole breast ultrasound using topographic watershed. *IEEE Trans. Med. Imaging* **33**, 1503–1511 (2014).
28. Mangan, A. P. & Whitaker, R. T. Partitioning 3d surface meshes using watershed segmentation. *IEEE Trans. Vis. Comput. Graph.* **5**, 308–321 (1999).
29. Zhang, J. *et al.* Fast and accurate automated recognition of the dominant cells from fecal images based on Faster R-CNN. *Sci. Rep.* **11**, 1–8 (2021).
30. Bochkovskiy, A., Wang, C.-Y. & Liao, H.-Y. M. Yolov4: Optimal speed and accuracy of object detection. *arXiv preprint* arXiv:2004.10934 (2020).
31. Moon, W. K. *et al.* Computer-aided tumor detection in automated breast ultrasound using a 3-d convolutional neural network. *Comput. Methods Programs Biomed.* **190**, 105360 (2020).
32. 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 Trans. Med. Imaging* **38**, 240–249 (2019).
33. Lei, Y. *et al.* Breast tumor segmentation in 3D automatic breast ultrasound using Mask scoring R-CNN. *Med. Phys.* **48**, 204–214 (2021).
34. Lei, Y. *et al.* Automatic tumor segmentation in 3D automated breast ultrasound using convolutional neural network. In *Medical Imaging 2021: Ultrasonic Imaging and Tomography*, vol. 11602, 276–281 (SPIE, 2021).
35. Zhang, J. *et al.* Application of convolution neural network algorithm based on multicenter abus images in breast lesion detection. *Front. Oncol.* **12**, 938413 (2022).
36. Jocher, G. *et al.* ultralytics/yolov5: v3.1—Bug Fixes and Performance Improvements (2020).
37. Malekmohammadi, A., Barekatrezaei, S., Kozegar, E. & Soryani, M. Mass detection in automated 3-d breast ultrasound using a patch bi-convlstm network. *Ultrasonics* **129**, 106891 (2023).
38. Zhou, Y. *et al.* 3D multi-view tumor detection in automated whole breast ultrasound using deep convolutional neural network. *Expert Syst. Appl.* **168**, 114410 (2021).
39. Inc., T. M. Matlab version: 9.12.0.1884302 (r2022a) (2022). https://www.mathworks.com.
40. D'Orsi, C. *2013 ACR BI-RADS Atlas: Breast Imaging Reporting and Data System* (American College of Radiology, 2014). https://books.google.co.kr/books?id=nhWSjwEACAAJ.
41. Lu, Y. *et al.* Identification of metastatic lymph nodes in MR imaging with faster region-based convolutional neural networks. *Cancer Res.* **78**, 5135–5143 (2018).
42. Li, H. *et al.* An improved deep learning approach for detection of thyroid papillary cancer in ultrasound images. *Sci. Rep.* **8**, 1–12 (2018).
43. Rosati, R. *et al.* Faster R-CNN approach for detection and quantification of DNA damage in comet assay images. *Comput. Biol. Med.* **123**, 103912 (2020).
44. Su, Y., Li, D. & Chen, X. Lung nodule detection based on Faster R-CNN framework. *Comput. Methods Programs Biomed.* **200**, 105866 (2021).
45. He, K., Zhang, X., Ren, S. & Sun, J. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **37**, 1904–1916 (2015).
46. Girshick, R., Donahue, J., Darrell, T. & Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. CVPR* **2014**, 580–587 (2014).
47. Lin, T.-Y. *et al.* Microsoft COCO: common objects in context. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13*, 740–755 (Springer, 2014).
48. Robbins, H. & Monro, S. A stochastic approximation method. *Ann. Math. Stat.* **22**, 400–407 (1951).
49. Sibson, R. Slink: An optimally efficient algorithm for the single-link cluster method. *Comput. J.* **16**, 30–34 (1973).
50. Jaccard, P. The distribution of the flora in the alpine zone 1. *New Phytol.* **11**, 37–50 (1912).
51. Deng, J. *et al.* Imagenet: A large-scale hierarchical image database. *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. CVPR* **2009**, 248–255 (2009).
52. Student. The probable error of a mean. *Biometrika* **6**, 1–25 (1908).
53. Demšar, J. Statistical comparisons of classifiers over multiple data sets. *J. Mach. Learn. Res.* **7**, 1–30 (2006).

Acknowledgements

Author contributions

K.O.: research and experimental planning, image preprocessing, deep learning system design, experiments, results analysis, manuscript writing, and manuscript revision; S.E.L. & E.K.K.: research and experimental planning, image grading, results analysis, manuscript writing, and manuscript revision.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to E.-K.K.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.