Weakly supervised Branch Network with Template Mask for Classifying Masses in 3D Automated Breast Ultrasound

Daekyung Kim[1,2] Chang-Mo Nam[2] Haesol Park[4] Mijung Jang[3] Kyong Joon Lee[2,3]
[1]Seoul National University [2]Monitor Corporation
[3]Seoul National University Bundang Hospital [4]Korea Institute of Science and Technology
sunnmoon137@snu.ac.kr

Abstract

Automated breast ultrasound (ABUS) is being rapidly utilized for screening and diagnosing breast cancer. Breast masses, including cancers shown in ABUS scans, often appear as irregular hypoechoic areas that are hard to distinguish from background shadings. We propose a novel branch network architecture incorporating segmentation information of masses in the training process. By providing the spatial attention effect, the branch network boosts the performance of existing neural network classifiers, helping to learn meaningful features around the mass. For the segmentation information, we leverage the existing radiology reports without additional labeling efforts. The reports should include the characteristics of breast masses, such as shape and orientation, and a template mask can be created in a rule-based manner. Experimental results show that the proposed branch network with a template mask significantly improves the performance of existing classifiers.

1. Introduction

Breast cancer is the second leading cause of cancer death in women worldwide [18]. Many studies have revealed that breast screening can discover cancer in its early stages reducing mortality [13, 17, 15]. Mammography has been the most widely used examination for breast screening, but it often misses cancer hidden in *dense* breast tissue that contains much fibrous or glandular tissue rather than fat [20]. Hand-held ultrasound (HHUS) imaging is a popular alternative covering dense breast tissue, but it suffers from low reproducibility depending on its operator. Recently, automated breast ultrasound (ABUS) has been introduced and is receiving favorable reviews [22]. While an HHUS makes partial two-dimensional breast scans that need to be simultaneously examined by an operator on site, an ABUS produces whole three-dimensional (3D) breast scans with a dedicated probe that can be asynchronously examined later. This has led to the active development of computer-aided diagnosis (CAD) systems for ABUS, that can assist interpreting physicians by reducing their workload and enhancing cancer detection performance [23, 8, 21].

The main task of CAD systems is to detect breast masses that can possibly grow into cancer. Breast masses on ABUS scans are usually shown as hypoechoic areas that can also appear due to a variety of causes such as fat, shadow, and anechoic tissue; thus many CAD systems erroneously detect these areas as suspicious breast masses, resulting in *false positives*. Even radiology professionals have difficulty distinguishing hypoechoic areas of breast masses from those of other causes [2, 9]. To develop competitive classifiers on ABUS, recent studies have employed convolutional neural networks (CNNs) that have been shown tremendous success in image classification tasks. Chiang et al. [1] searched for volumes of interest with a fast sliding window, then applied a simple three-dimensional extension of a 2-D CNN to compute probabilities of tumor candidates. Moon et el. [14] incorporated a focal loss and ensemble learning into their 3D CNN model to resolve a data imbalance issue. Although those methods sensitively detect breast masses, they yield an average of more than four false positives per scan. Simple modifications of the existing CNNs may not be sufficient to analyze the three-dimensional context of breast masses on ABUS scans.

Training a CNN requires many sample images that preferably contain clear features of target objects. However, breast masses shown in ABUS scans often appear as irregularly shaded blobs, and each blob itself is hardly distinguishable from its background shadings. Considering these characteristics, we empirically found that training with a segmentation *mask* on a mass helps improve the classification performance, probably because the low-level features on the blobs are activated on the masses rather than on the backgrounds. The second row of Fig. 1 shows class activation maps (CAMs) for mass classification using the existing DenseNet [7]. The activated network weights are widely distributed across the background blobs and are not con-

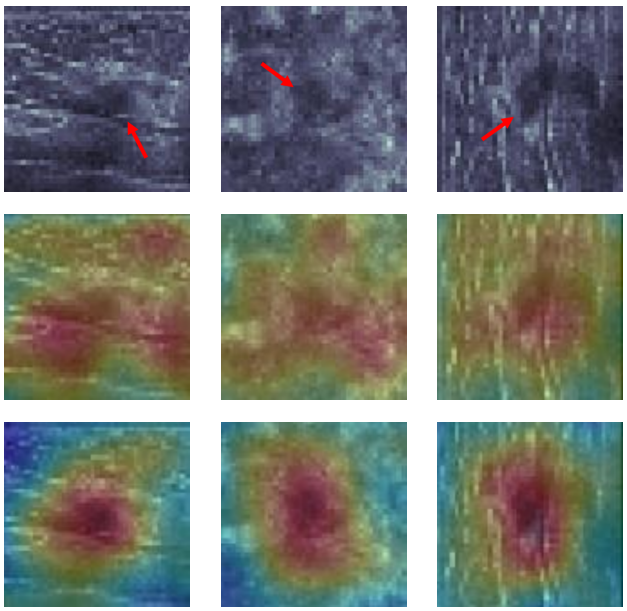

Figure 1: Class activation maps (CAMs) for sample mass images. The first row shows slice images of the breast mass in different views: transverse, coronal, and sagittal. The second row shows the CAM results from the existing DenseNet classifier. The third row shows CAM results from the proposed method. An activation with a high weight is shown in red color.

centrated on the masses. In contrast, our modified classifier with segmentation masks intensively activates the weights on the mass blobs, as illustrated in the third row. To integrate the mask information, we attached an additional network (called the *mask branch* network) to the middle of a classification network. Since our branch network is independently applicable, the proposed method can be modified from any existing state-of-the-art CNN model.

Another practical issue lies in the difficulty of generating segmentation masks. Aside from the high cost of professional tasks within a three-dimensional space, the exact boundary is not clear even for medical experts. We propose to employ a *template mask* with pre-defined shapes instead (e.g., circle, oval), with minimal parameters such as diameter and direction. This information is typically recorded in radiology reports, thus additional labeling effort is rarely required.

2. Method

The overall architecture of the proposed method is illustrated in Fig. 2. Our network architecture consists of two parallel networks: the main network and the mask branch network. The main network works as a backbone CNN classifier. In the middle of the main network, the *mask branch network* bifurcates off to compute the mask loss by comparing it with the template mask generated from a radiology report. The main loss and the mask loss of each network are integrated into the total loss to be optimized in the training process.

2.1. Main Network

The main network implements a binary classifier determining the existence of suspicious breast lesions. It is composed of four dense blocks with transition layers based on DenseNet-BC-121 [7]. A transition layer is placed between two consecutive dense blocks to perform downsampling, batch normalization, $1 \times 1 \times 1$ convolution, and average pooling. The global average pooling (GAP) layer is connected at the end of the last dense block, followed by the fully connected (FC) layer and the softmax activation layer sequentially. The main loss can be defined as a cross entropy loss of input samples, shown as follows:

$$L_{main} = \frac{1}{N}\sum_{i}^{N}[(1-y_i)\log(1-p_i) + y_i \log p_i], \quad (1)$$

where i means the index of an input sample, N is the number of samples, $y_i \in \{0, 1\}$ is the ground-truth label, and p_i is the probability estimating whether the input sample contains a mass or not.

2.2. Mask Branch Network

The concept of branch architecture [19, 3] is to train multiple tasks that take advantage of the interaction between different tasks. The mask branch aims to integrate spatial information into the main network, which helps the main network extract meaningful features from the area around the mass.

The mask branch is designed to start at the $1 \times 1 \times 1$ convolutional layer of the second transition layer of the main branch, and to generate a voxel-level probability map (branch output) that estimates the presence of target lesions. We may assume that this branch output simulates a segmentation mask.

We denote $\mathbf{F} \in \mathbb{R}^{C\times H\times W\times D}$ as a midlayer CNN feature extracted from the second transition layer, where C, H, W and D are the number of channels, height, width and depth of the feature map, respectively.

The *branch output* $\mathbf{B}$ can be formulated as:

$$\mathbf{B} = softmax(f(\sigma(\mathbf{F}), w)), \quad (2)$$

where $f(\cdot,\cdot)$ denotes a 3D convolution function, σ represents the activation function, $\mathbf{B} \in \mathbb{R}^{H\times W\times D\times 2}$ means the generated *branch output* map, and $w \in \mathbb{R}^{2\times C\times 1\times 1\times 1}$ indicates convolutional parameters. As a result, branch output is a voxel-level binary probability map of the target lesion and background.

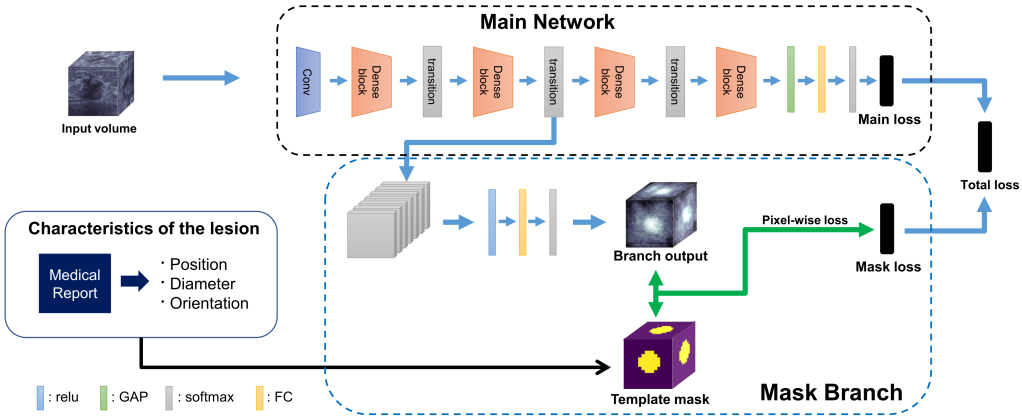

Figure 2: The architecture of our proposed network: mask branch network

To integrate the spatial attention, the branch output is compared to the template mask by computing a pixelwise cross-entropy loss. The definition of mask loss is shown as follows:

$$L_{mask} = \frac{1}{NM}\sum_{i}^{N}\sum_{j}^{M}[(1-z_{i,j})\log(1-q_{i,j}) + z_{i,j}\log q_{i,j}], \tag{3}$$

where j is the index of a voxel, and M is the number of voxels in the branch output. $z_{i,j} \in \{0,1\}$ is the label value in the template mask and $q_{i,j}$ is the probability value of the branch output, at the j_{th} voxel of the i_{th} sample, respectively.

The total loss to be optimized is defined as the combination of the main loss and the mask loss using the uncertainty loss weighting method [10]:

$$L_{total} = \frac{1}{\sigma_1^2}L_{main} + \frac{1}{\sigma_2^2}L_{mask} + \log\sigma_1 + \log\sigma_2, \tag{4}$$

where σ_1 and σ_2 are trainable variables that adaptively learn the relative weight of L_{main} and L_{mask} and regulate the balance of the losses.

mask branch network is used only for the training process to focus on the region of interest and not for the testing process. Thus, the proposed network requires no additional parameters in the inference process.

2.3. Template Mask

For a segmentation mask, a pixelwise segmentation map fully annotated by a clinical expert can be an ideal mask; however, manually tracking the boundaries of a three-dimensional mass requires considerable time and effort from the expert. We propose to utilize the information that already exists in the radiology report instead of manually labeling the ground truth.

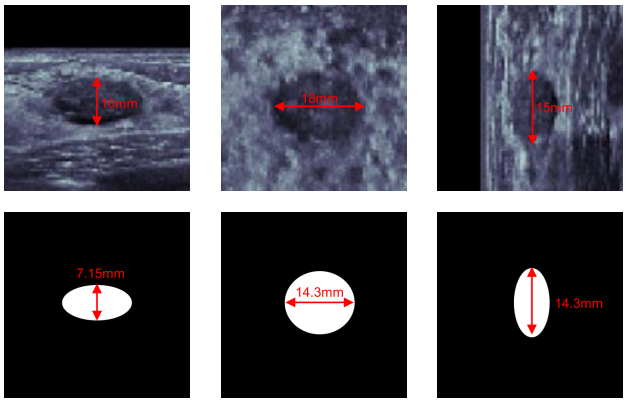

Figure 3: Example of template mask with 14.3mm average diameter and parallel orientation. **Top row:** transverse, coronal, and sagittal images of breast cancer on ABUS with diameters annotated on each axis. **Bottom row:** template masks corresponding to the top row images.

The location, size and category of suspicious lesions are usually written in radiology reports during normal diagnostics. By annotating the three axes of the lesion, its location and size are recorded in the form of center coordinates and average diameter. For categories, the BI-RADS (Breast Imaging Reporting and Data System) criteria [16] are the most widely used, which include several visual elements such as shape, orientation, margins, echo patterns, and posterior features.

We employ the orientation features, indicating that the direction of the long axis relative to the breast skin is *parallel* or not: If the orientation is parallel, the lesion is assumed to be a complete sphere, otherwise it is assumed to be an ellipsoid whose diameter perpendicular to the skin is halved.

Fig. 4 shows the process of creating a *template mask* by each step. Similar to the general segmentation map, *tem-*

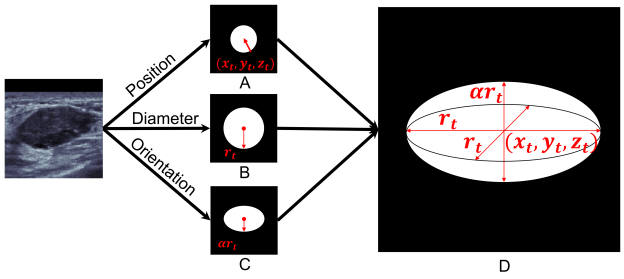

Figure 4: How to make *template mask* by utilizing characteristics of target lesions

plate mask is a binary label map that consists of positive and negative areas. The positive area is created by the rule utilizing the position, diameter, and *orientation* of breast lesions. First, the *template mask* begins with a sphere with the centroid on the (x_t, y_t, z_t) center position of the lesion recorded in the radiology report(Fig. 4A). In addition to the center position, the radius of the sphere is defined as the average radius r_t of the lesion.(Fig. 4B). Finally, if the lesion has a '*parallel*' property of *orientation*, the diameter on perpendicular to the skin shrinks to $r_t/2$ changing the shape of the positive area from a sphere to an ellipsoid(Fig. 4C). As a result, the equation for *template mask* can be written as below:

With the center coordinate (x_t, y_t, z_t), diameter d_t and orientation of the lesion, we can define a template mask $\mathcal{T}(x, y, z)$, a simplified segmentation label.

$$\mathcal{T}(x,y,z) = \begin{cases} 1 & \text{if } \frac{(x-x_t)^2}{d_t} + \frac{(y-y_t)^2}{\alpha d_t} + \frac{(z-z_t)^2}{d_t} \leq 1, \\ 0 & \text{else} \end{cases} \quad (5)$$

where α is 0.5 if the orientation is parallel otherwise, $\alpha = 1$.

An example of template mask is shown in Fig. 3. The top row shows transverse, coronal, and sagittal views of the center of breast mass. The orientation is parallel and the average diameter is $14.3mm$, with diameters of each axis are $10mm$, $18mm$, $15mm$ respectively. As seen, the template mask generally fits into the actual mass area.

3. Experiments

To the best of our knowledge, no public dataset is available for ABUS research, thus we built our own dataset in a tertiary hospital. ABUS images used in this study were acquired from ABUS systems (Invenia ABUS, Automated Breast Ultrasound System; GE Healthcare, Sunnyvale, CA, USA). For each breast, three volumes were obtained: the central volume, the lateral volume, and the medial volume. The institutional review board approved this study and waived informed consent, considering the retrospective study design and the use of anonymized patient data.

A total of 363 patients who underwent ABUS from May 2017 to October 2019 were included. ABUS images of 286 patients presented 434 mass lesions categorized as $C2$ or above by BI-RADS, while other 77 images showed no mass lesions. We randomly divided the mass lesions: 304 lesions in the training set, 50 lesions in the validation set, 80 lesions in the test set. The dataset also included $3,907$ nonmass lesions that are randomly cropped from the 77 normal patients' images, and randomly divided into $3,777$ lesions for the training set, and 50 and 80 lesions for the validation and test sets, respectively.

All center coordinates and diameters of masses were obtained from radiology reports annotated by experienced radiologists. We rescaled the original volume images to have voxel sizes of $0.3mm$ to $2.0mm$ depending on the mass size, and then we cropped the network input volumes at the center coordinate of the mass with the size of $48 \times 32 \times 48$ voxels. Additionally, the 3-dimensional patches used for training are rotated three times with rotation angles of $90\,^\circ$, $180\,^\circ$, and $270\,^\circ$ for augmentation.

3.1. Evaluation metrics.

The sensitivity and specificity are the percentage of positive and negative results that are correctly identified. which is defined as:

$$Sensitivity(Se) = \frac{|TP|}{|TP| + |FN|} \quad (6)$$

$$Specificity(Sp) = \frac{|TN|}{|TN| + |FP|} \quad (7)$$

where TP and TN are *true positive* and *true negative* and likewise FP and FN are *false negative* and *false positive*. The AUC was calculated by using receiver operating characteristic (ROC) analysis on the test set.

All networks in this study were implemented with the TensorFlow 1.12 library and were trained on Nvidia 2080-Ti on an Ubuntu 18.04 system. The weights of the networks adopted the Xavier uniform initialization [4]. The batch size was set to 48 and the total loss was optimized by the RMSprop optimizer.

The classification performance was evaluated by measuring the sensitivity (Se), specificity (Sp), and area under the curve (AUC) of receiver operating characteristics (ROC).

3.2. Branch network evaluation

To evaluate the effect of the proposed mask branch network (MBN), we tested the main network as a baseline, and compared the results from the main network with MBN. We employed two networks as the main network: DenseNet-BC [7] and ResNetV2-101[6]. As shown in Table 1, the proposed networks with MBN outperform the baselines in every performance measure, in both kinds of main networks.

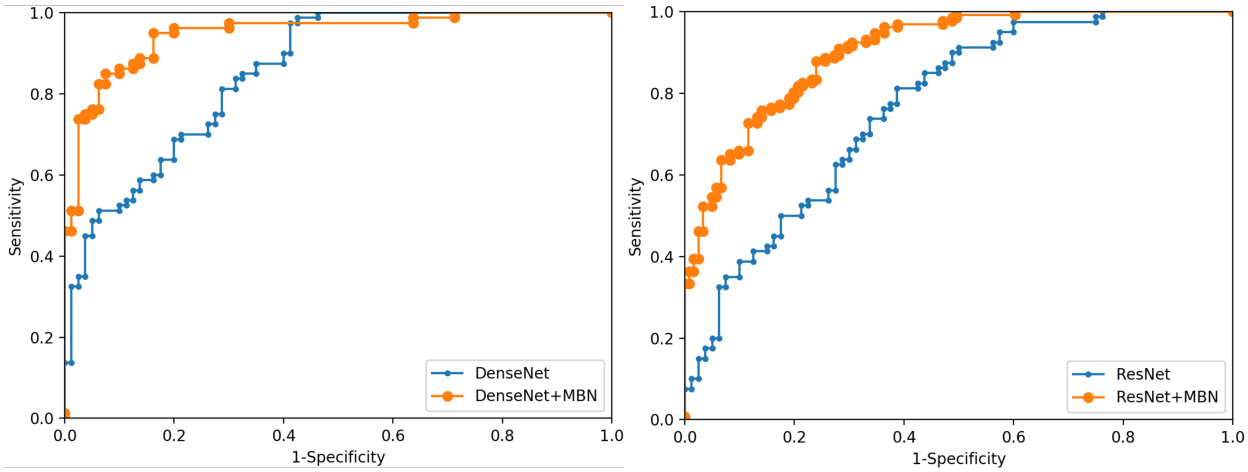

Figure 5: ROC curves of *mask branch network*(MBN) with DenseNet and ResNet.

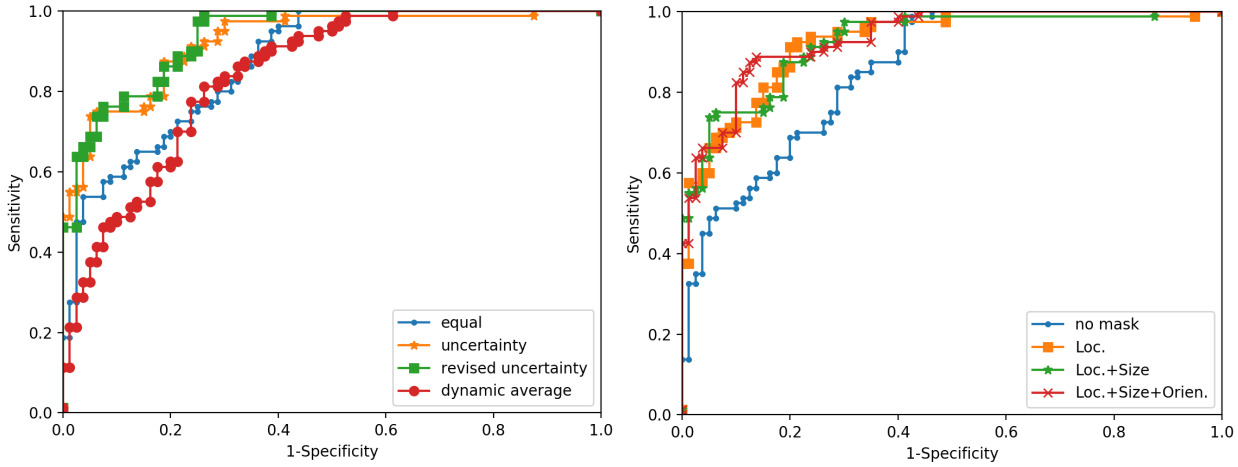

Figure 6: ROC curves. (Left): Models using various ways to weigh multiple losses; (Right): Cases utilizing different characteristics for breast lesions.

Table 1: Performance comparison for applying the mask branch network (MBN)

	Se	**Sp**	**AUC**
DenseNet	56.3%	86.3%	0.855
DenseNet + MBN	**75.0%**	**92.5%**	**0.9252**
ResNet	65.0%	70.0%	0.768
ResNet + MBN	**68.8%**	**77.5%**	**0.827**

Fig. 7 qualitatively visualizes the spatial attention effect of the proposed mask branch by plotting the grad-CAM (class activation map) visualization. The first row shows sample mass images with red rectangles indicating the masses. The CAM of DenseNet (second row) do not appropriately activate areas where the masses present. In contrast, the proposed method (third row) shows relatively clear activation on the target areas; thus the main network is expected to learn to exploit and aggregate features in the target area.

3.3. Comparison of loss weighting strategies

We compared weighting loss methods with mask loss. Various methods have been used to efficiently combine multiple losses in multitask learning (MTL)[5]. The most straightforward method is uniform weighting: the losses are simply added together to produce a single scalar loss value. Dynamic optimization techniques are also important in MTL to optimize the set of possibly contrasting losses or gradients because conflicting gradient problems

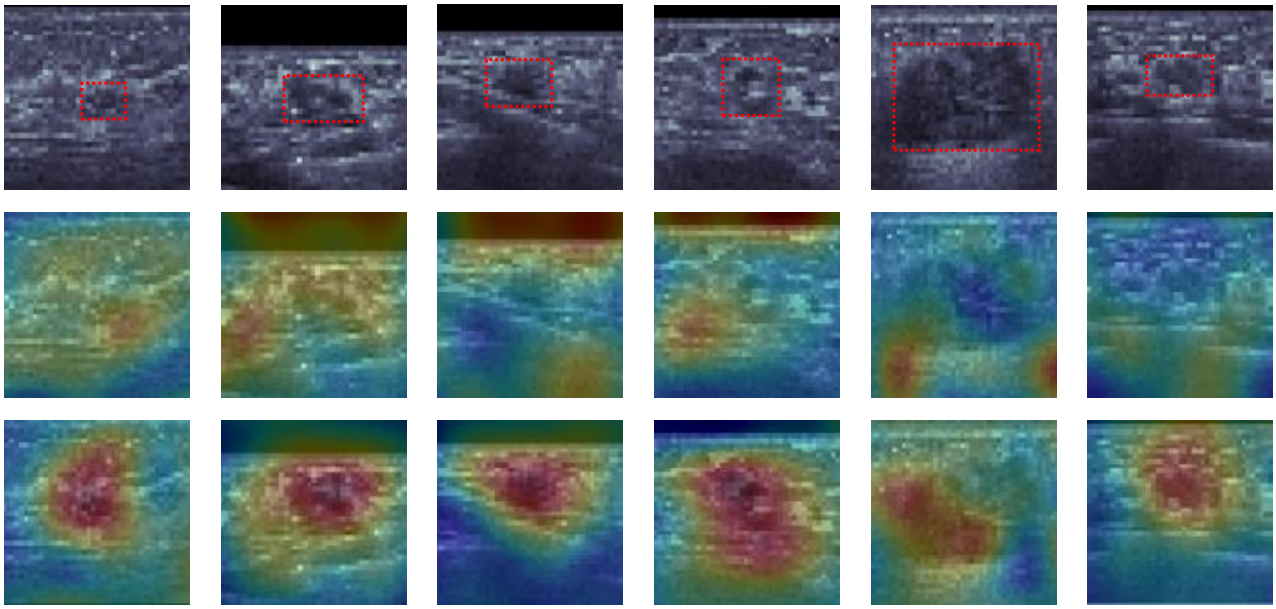

Figure 7: Class activation maps (CAMs) to show the effect of the mask branch network (MBN). **Top row:** Breast mass samples on ABUS images marked by red boxes. **Middle row:** CAMs of DenseNet. **Bottom row:** CAMs of DenseNet with MBN

Table 2: Experiments on ways to weigh the main loss and the mask loss. Metrics ranked with 1st place are in bold.

	Se	Sp	AUC
Equal	65.00%	82.50%	0.8720
Uncertainty	**75.00%**	92.50%	0.9252
Revised Uncertainty	71.25%	**93.75%**	**0.9369**
Dynamic Average	70.00%	77.50%	0.8416

Table 3: Comparison on features sets used in creating the template masks

	Se	Sp	AUC
No mask	56.3%	86.3%	0.855
Loc.	67.5%	90.0%	0.908
Loc.+Size	72.5%	90.0%	0.921
Loc.+Size+Orien.	**75.0%**	**92.5%**	**0.925**

can degrade model performance. Kendall et al.[10] proposed an uncertainty-based weighting approach and applied it to a CNN. [11] adapted the regularization term in Kendall's uncertainty-based method. A dynamic weight average (DWA) was proposed by [12]. We adapted these methods to the proposed method with MBN-V1 and exactly the same experimental setting and compared the performance. The results are shown in Table 2, which shows that the uncertainty weighting and revised uncertainty weighting achieve higher performance compared to other settings. The results do not show many differences between uncertainty weight and revised uncertainty weight, but their performance is considerably better than the others. Considering this result, we applied uncertainty weighting to our method for the final result.

3.4. Template mask evaluation

To evaluate the effect of mass features available in radiology reports, we compared the performance of template masks utilizing various feature sets and provided the results in table 3. Even utilizing location only (with a fixed sphere of $20mm$ diameter) helps to improve all performance metrics compared to the baseline (DenseNet without MBN; no mask) The size (diameter) also enhanced the metrics and the combination of all features yielded the best performance as seen.

4. Conclusion

This study introduced our novel branch network incorporating attention information to improve the performance of CNN for classifying masses on ABUS images. We utilized the characteristics of breast mass recorded in radiology reports to generate a simplified mask that required no

additional labor for segmentation. The proposed MBN can be attached to existing networks and has been shown to boost performance. In future work, we will test variations of branch network to find the optimal architecture. Also, the method will be applied to more challenging problems on ABUS such as cancer (malignancy)

Acknowledgement

This work was supported by the Industrial Strategic technology development program (10072064) funded by the Ministry of Trade, Industry and Energy (MI, Korea) and by grant (no. 13-2019-006) from the SNUBH Research Fund.

References

[1] Tsung-Chen Chiang, Yao-Sian Huang, Rong-Tai Chen, Chiun-Sheng Huang, and Ruey-Feng Chang. Tumor detection in automated breast ultrasound using 3-d cnn and prioritized candidate aggregation. *IEEE transactions on medical imaging*, 38(1):240–249, 2018.

[2] Pavel Crystal, Selwyn D Strano, Semyon Shcharynski, and Michael J Koretz. Using sonography to screen women with mammographically dense breasts. *American Journal of Roentgenology*, 181(1):177–182, 2003.

[3] Hiroshi Fukui, Tsubasa Hirakawa, Takayoshi Yamashita, and Hironobu Fujiyoshi. Attention branch network: Learning of attention mechanism for visual explanation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10705–10714, 2019.

[4] Xavier Glorot and Yoshua Bengio. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, pages 249–256, 2010.

[5] Ting Gong, Tyler Lee, Cory Stephenson, Venkata Renduchintala, Suchismita Padhy, Anthony Ndirango, Gokce Keskin, and Oguz H Elibol. A comparison of loss weighting strategies for multi task learning in deep neural networks. *IEEE Access*, 7:141627–141632, 2019.

[6] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Identity mappings in deep residual networks. In *European conference on computer vision*, pages 630–645. Springer, 2016.

[7] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708, 2017.

[8] Yulei Jiang, Marc F Inciardi, Alexandra V Edwards, and John Papaioannou. Interpretation time using a concurrent-read computer-aided detection system for automated breast ultrasound in breast cancer screening of women with dense breast tissue. *American Journal of Roentgenology*, 211(2):452–461, 2018.

[9] Kevin M Kelly, Judy Dean, W Scott Comulada, and Sung-Jae Lee. Breast cancer detection using automated whole breast ultrasound and mammography in radiographically dense breasts. *European radiology*, 20(3):734–742, 2010.

[10] Alex Kendall, Yarin Gal, and Roberto Cipolla. Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7482–7491, 2018.

[11] Lukas Liebel and Marco Körner. Auxiliary tasks in multi-task learning. *arXiv preprint arXiv:1805.06334*, 2018.

[12] Shikun Liu, Edward Johns, and Andrew J Davison. End-to-end multi-task learning with attention. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1871–1880, 2019.

[13] Anthony B Miller, Claus Wall, Cornelia J Baines, Ping Sun, Teresa To, and Steven A Narod. Twenty five year follow-up for breast cancer incidence and mortality of the canadian national breast screening study: randomised screening trial. *Bmj*, 348, 2014.

[14] Woo Kyung Moon, Yao-Sian Huang, Chin-Hua Hsu, Ting-Yin Chang Chien, Jung Min Chang, Su Hyun Lee, Chiun-Sheng Huang, and Ruey-Feng Chang. Computer-aided tumor detection in automated breast ultrasound using a 3-d convolutional neural network. *Computer Methods and Programs in Biomedicine*, 190:105360, 2020.

[15] Advisory Committee on Breast Cancer Screening. Screening for breast cancer in england: past and future. *Journal of medical screening*, 13(2):59–61, 2006.

[16] Susan G Orel, Nicole Kay, Carol Reynolds, and Daniel C Sullivan. Bi-rads categorization as a predictor of malignancy. *Radiology*, 211(3):845–850, 1999.

[17] Debbie Saslow, Carla Boetes, Wylie Burke, Steven Harms, Martin O Leach, Constance D Lehman, Elizabeth Morris, Etta Pisano, Mitchell Schnall, Stephen Sener, et al. American cancer society guidelines for breast screening with mri as an adjunct to mammography. *CA: a cancer journal for clinicians*, 57(2):75–89, 2007.

[18] Rebecca L Siegel, Kimberly D Miller, and Ahmedin Jemal. Cancer statistics, 2019. *CA: a cancer journal for clinicians*, 69(1):7–34, 2019.

[19] Surat Teerapittayanon, Bradley McDanel, and Hsiang-Tsung Kung. Branchynet: Fast inference via early exiting from deep neural networks. In *2016 23rd International Conference on Pattern Recognition (ICPR)*, pages 2464–2469. IEEE, 2016.

[20] M Van Goethem, K Schelfout, L Dijckmans, JC Van Der Auwera, J Weyler, I Verslegers, I Biltjes, and A De Schepper. Mr mammography in the pre-operative staging of breast cancer in patients with dense breast tissue: comparison with mammography and ultrasound. *European radiology*, 14(5):809–816, 2004.

[21] Jan CM van Zelst, Tao Tan, Paola Clauser, Angels Domingo, Monique D Dorrius, Daniel Drieling, Michael Golatta, Francisca Gras, Mathijn de Jong, Ruud Pijnappel, et al. Dedicated computer-aided detection software for automated 3d breast ultrasound; an efficient tool for the radiologist in supplemental screening of women with dense breasts. *European radiology*, 28(7):2996–3006, 2018.

[22] Athina Vourtsis and Aspasia Kachulis. The performance of 3d abus versus hhus in the visualisation and bi-rads charac-

terisation of breast lesions in a large cohort of 1,886 women. *European radiology*, 28(2):592–601, 2018.

[23] Shanling Yang, Xican Gao, Liwen Liu, Rui Shu, Jingru Yan, Ge Zhang, Yao Xiao, Yan Ju, Ni Zhao, and Hongping Song. Performance and reading time of automated breast us with or without computer-aided detection. *Radiology*, 292(3):540–549, 2019.