Journal Pre-proof

MommiNet-v2: Mammographic Multi-View Mass Identification Networks

Zhicheng Yang, Zhenjie Cao, Yanbo Zhang, Yuxing Tang, Xiaohui Lin, Rushan Ouyang, Mingxiang Wu, Mei Han, Jing Xiao, Lingyun Huang, Shibin Wu, Peng Chang, Jie Ma

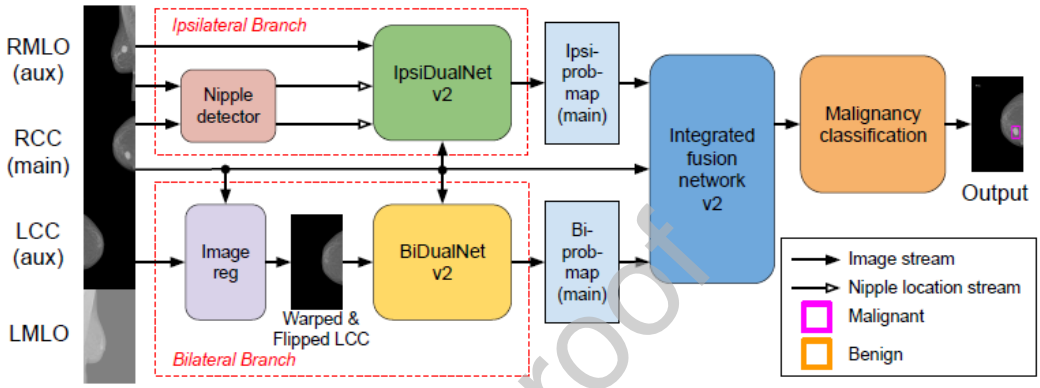
PII: S1361-8415(21)00249-8
DOI: https://doi.org/10.1016/j.media.2021.102204
Reference: MEDIMA 102204

To appear in: *Medical Image Analysis*

Received date: 25 December 2020
Revised date: 12 April 2021
Accepted date: 8 June 2021

Please cite this article as: Zhicheng Yang, Zhenjie Cao, Yanbo Zhang, Yuxing Tang, Xiaohui Lin, Rushan Ouyang, Mingxiang Wu, Mei Han, Jing Xiao, Lingyun Huang, Shibin Wu, Peng Chang, Jie Ma, MommiNet-v2: Mammographic Multi-View Mass Identification Networks, *Medical Image Analysis* (2021), doi: https://doi.org/10.1016/j.media.2021.102204

GRAPHICAL ABSTRACT



HIGHLIGHTS

- We further improve MommiNet, the first tri-view DNN architecture to perform joint ipsilateral and bilateral analysis for mass detection, and present MommiNet-v2 to fully aggregate information from the high-resolution representations of all views with improved mass detection performance.
- A multi-task learning scheme to incorporate the malignancy information from both biopsy test and BI-RADS categories, for improved mass malignancy classification.
- SOTA Free-Response Operating Characteristic (FROC) mass detection performance achieved on the entire DDSM and our in-house datasets.

MommiNet-v2: Mammographic Multi-View Mass Identification Networks

Zhicheng Yang[a,1], Zhenjie Cao[a,1], Yanbo Zhang[a], Yuxing Tang[a], Xiaohui Lin[c], Rushan Ouyang[c], Mingxiang Wu[c], Mei Han[a], Jing Xiao[b], Lingyun Huang[b], Shibin Wu[b], Peng Chang[a,*], Jie Ma[c,*]

[a]*PAII Inc., Palo Alto, CA 94306, USA*
[b]*Ping An Technology Co., Ltd., Shenzhen, 518000, China*
[c]*Department of Radiology, Shenzhen Peoples Hospital (The Second Clinical Medical College, Jinan University) Shenzhen 518020, Guangdong, China*

Abstract

Many existing approaches for mammogram analysis are based on single view. Some recent DNN-based multi-view approaches can perform either bilateral or ipsilateral analysis, while in practice, radiologists use both to achieve the best clinical outcome. MommiNet is the first DNN-based tri-view mass identification approach, which can simultaneously perform bilateral and ipsilateral analysis of mammographic images, and in turn, can fully emulate the radiologists' reading practice. In this paper, we present MommiNet-v2, with improved network architecture and performance. Novel high-resolution network (HRNet)-based architectures are proposed to learn the symmetry and geometry constraints, to fully aggregate the information from all views for accurate mass detection. A multi-task learning scheme is adopted to incorporate both Breast Imaging-Reporting and Data System (BI-RADS) and biopsy information to train a mass malignancy classification network. Extensive experiments have been conducted on the public DDSM (Digital Database for Screening Mammography) dataset and our in-house dataset, and state-of-the-art results have been achieved in terms of mass detection accuracy. Satisfactory mass malignancy classification result has also been obtained on our in-house dataset.

Keywords: Mammogram, Deep learning, Multi-view, Mass, Malignancy

*Corresponding authors. Email addresses: pengchang@gmail.com, majie688@hotmail.com
[1]These authors equally contributed to this work.

1. Introduction

Mammography is widely used as a cost-effective early detection method for breast cancer, the most common cancer in women worldwide and the second leading cause of cancer death for women in the US. With about 39 million mammograms performed annually in the US alone, Computer-Aided Diagnosis (CAD) systems have the promise to help radiologists improve the overall efficiency and accuracy for breast cancer diagnosis. Significant progress has recently been made in the performance of CAD systems, especially with the advance of DNN-based methods. Nonetheless, mammographic abnormality detection and malignancy classification remain challenging, largely due to the high accuracy requirement set by the clinical practice.

There are generally two categories of approaches toward a CAD system for mammograms: multi-stage and end-to-end. The multi-stage approaches follow the diagnostic routine of a radiologist, by dividing the whole process into several stages, such as identifying all the lesion regions, classifying the malignancy for each lesion, and reporting the overall cancer risk. On the contrary, the end-to-end approaches take the mammographic images as input, and directly output the cancer risk at the image or patient level, bypassing the lesion level output. Some recent results show that the end-to-end systems have the potential to outperform the radiologists in certain circumstances (McKinney et al., 2020; Wu et al., 2019; Akselrod-Ballin et al., 2019), nonetheless it is still a far-fetched goal to replace the radiologists in the near future. Our collaboration with the radiologists in the clinical experiments indicates that the radiologists often prefer to have the lesion level output from the CAD system, to make the final diagnosis decisions. Therefore we take a multi-stage (Guan et al., 2020) approach, and in this paper, we focus on mass detection and malignancy classification in mammograms.

A standard mammography screening procedure acquires two low-dose X-ray projection views for each breast, a craniocaudal (CC) view and a mediolateral oblique (MLO) view. Radiologists routinely use all views in breast cancer diagnosis. The ipsilateral analysis refers to the diagnosis based on the CC and MLO views of the same breast, while the bilateral analysis combines the findings from the same views

of the two breasts. For example, the radiologists may cross-check the lesion locations through the ipsilateral analysis, and use the symmetry information from the bilateral analysis to improve the decision accuracy. Many previous approaches on mammographic lesion detection focus on one view (Li et al., 2018; Agarwal et al., 2019; Zhang et al., 2019; Xi et al., 2018; Cao et al., 2019a,b; Li et al., 2019), therefore unable to capture the rich information from the multiple view analysis. Recently several DNN-based dual-view approaches have been proposed, performing either ipsilateral or bilateral analysis (Carneiro et al., 2017; Perek et al., 2018; Ren et al., 2019; Diniz et al., 2018; Liu et al., 2019; Li et al., 2020). In our previous work (Yang et al., 2020b), we have proposed MommiNet (MammOgraphic Multi-view Mass Identification NETworks [**maa**·mee·net]), the first DNN-based architecture to perform tri-view based mass detection. In this work, we further improve this model by incorporating the recently proposed High-Resolution Network (HRNet) as the backbone to to preserve high-resolution feature representations through the network (Wang et al., 2020c). Moreover, we also integrate a classification module to classify the malignancy of the detected masses. High-resolution feature representations are critical for the accuracy of both mass detection and malignancy classification in mammograms, since the detailed texture information of lesions are well-preserved. This upgraded version is denoted as MommiNet-v2.

BI-RADS (Breast Imaging-Reporting and Data System) (American College of Radiology, 2013) and biopsies are commonly used to assess the cancer risk of breast lesions. A radiologist typically assigns each lesion/breast a BI-RADS category from 0 to 6 in a diagnostic report after interpreting a mammogram. A biopsy can be ordered to confirm the malignancy for lesions with high BI-RADS levels, and is considered as the gold standard. Most previous studies in the literature treat the mass malignancy classification as a binary problem, and the mass malignancy labels are obtained from biopsy results. However, since biopsy is an invasive operation, only the patients with high risk will take this further test, while BI-RADS information is more widely available to indicate lesion malignancy. To take advantage of the BI-RADS information, in this work we adopt a multi-task learning framework (Caruana, 1997) for the mass malignancy classification to combine BI-RADS categories and biopsy results.

In all, we present a two-stage system for mass detection and malignancy classification. Our main contributions include:

- We further improve MommiNet, the first tri-view DNN architecture to perform joint ipsilateral and bilateral analysis, and present MommiNet-v2, which can fully aggregate information from high-resolution representations of all views with improved mass detection performance.
- A multi-task learning scheme to incorporate the malignancy information from both biopsies and BI-RADS categories, for improved mass malignancy classification.
- State-of-the-art (SOTA) Free-Response Operating Characteristic (FROC) mass detection performance on the entire DDSM and our in-house datasets.

A preliminary version of this work has appeared in (Yang et al., 2020b). In this paper, we further improve the mass detection with better model structures (high-resolution models), and provide more technical details of our method. Furthermore, we propose a malignancy classifier to output the BI-RADS category for each detected mass, along with a newly proposed multi-task learning framework to incorporate both BI-RADS and biopsy information from the training data. Extended experiment results with ablation studies are also included.

The rest of the paper is organized as follows. We review the related work in Section 2. In Section 3, we introduce the architecture of our method and loss functions. We present our experiment results and comparisons in Section 4, and discuss the results and limitations in Section 5. In Section 6, we conclude and describe future directions.

2. Related Work

2.1. Mass detection

Deep learning has been used to detect mass in mammograms, and most of the methods use a single image for detection (Li et al., 2018; Agarwal et al., 2019; Zhang et al., 2019; Xi et al., 2018; Cao et al., 2019a,b; Li et al., 2019). Recently, multi-view based approaches are attracting an increasing attention. In (Diniz et al., 2018;

Liu et al., 2019; Li et al., 2020), bilateral analysis has been incorporated in DNN-based approaches. Some other DNN-based methods consider information of ipsilateral mammograms (Carneiro et al., 2017; Perek et al., 2018; Ren et al., 2019). However, most of these approaches do not model the geometry relation across views explicitly. In (Ma et al., 2019), a cross-view relation network is added to the Siamese Networks for mass detection. However, this approach uses the same geometric features and embedding for the relation network as in (Hu et al., 2017), which was designed for single view object detection. In (Liu et al., 2020), a Bipartite Graph Convolutional Network is applied to detect masses, which considers spatial information of nodes from ipsilateral mammograms.

In this work, we perform the ipsilateral and bilateral analysis simultaneously using the specifically designed detection network, which is a Faster-RCNN (Ren et al., 2015) variant with Siamese input module, and the segmentation network, which is an HRNet (Wang et al., 2020c) variant with Siamese input module, respectively. Unlike (Ma et al., 2019), our relation network is explicitly designed to encode the mass-to-nipple distance for the ipsilateral analysis, in tandem with a DNN-based nipple detector. The mass-to-nipple distance has been considered in previous work (Sahiner et al., 2006), while our approach is the first one to explicitly embed this prior knowledge into a DNN architecture.

2.2. Mass malignancy classification

Different end-to-end malignancy classification approaches have been developed in at the image, breast, and patient level for breast cancer screening (McKinney et al., 2020; Wu et al., 2019; Akselrod-Ballin et al., 2019). The biopsy result can serve as the ground truth for mammographic images with a biopsy record. In addition, normal mammographic images can be obtained from patients who do not develop breast cancer in the following 12-24 months (Akselrod-Ballin et al., 2019). Wang et al. (Wang et al., 2020b) propose a Cycle-GAN (Cycle-Consistent Generative Adversarial Networks)-based model (Zhu et al., 2017) that uses bilateral symmetric prior and "healthy" image generation mechanisms to boost mammogram malignancy classification. The motivation behind the symmetric prior is that a lesion present on one side of the breasts

rarely appears in the corresponding area on the other side. A bilateral cycle-consistency mechanism is proposed and contralateral mammograms are used as references to generate the healthy version of target features to help find the abnormal features. In a later work (Wang et al., 2020a), they extend the model by considering both bilateral and epsilateral views as in (Yang et al., 2020b). The end-to-end classification approaches do not require the annotated lesion types and locations. Therefore, it is possible to obtain mammograms with image-level malignancy information at a large scale, which is essential for boosting the classification performance of a deep neural network model. However, the end-to-end methods are designed for image-level malignancy classification, and do not provide detailed lesion level output. Although the generated heat maps could highlight the potential regions of the malignant lesions, they generally cannot output the lesion types or the precise locations. In other words, the image-level classification can provide limited information to the radiologists in terms of identifying individual lesions.

Multi-stage approaches focus on lesion detection and malignancy classification for each lesion type, such as calcification and mass. Since mass detection is still a challenging problem, some previous mass malignancy classification methods have been developed with mass data labeled by human experts (Rangayyan et al., 1997; Wang et al., 2009; Pedro et al., 2019; Jiao et al., 2016; Wang et al., 2018). Most of them applied hand-crafted features or traditional machine learning methods (Rangayyan et al., 1997; Wang et al., 2009), and the classification performance is not optimal. Recently, some deep learning based techniques have resulted in improved classification performance for mass malignancy (Jiao et al., 2016; Wang et al., 2018).

A few integrated mass detection and classification systems have also been developed. Generally, most of these integrated systems implement the detection and classification with a cascaded or multi-stage manner (Al-Masni et al., 2018; Al-Antari et al., 2018; Abdelhafiz et al., 2019; Dhungel et al., 2017). In Dhungel et al. (2017), users are required to manually reject false positive mass regions. This user intervention sets a barrier in practical applications. By contrast, our proposed strategy and most cascaded-based methods (Al-Masni et al., 2018; Al-Antari et al., 2018; Abdelhafiz et al., 2019) do not have this limitation.

Unlike the aforementioned methods, our work adopts the multi-task learning framework and incorporates both the BI-RADS category information and the biopsy information when available for each mass during the training, which leads to a larger training dataset and much improved classification performance. Compared to the binary malignancy result from biopsy test, the multi-level BI-RADS categories provide more detailed information about the malignancy likelihood, which is more informative for training a deep neural network based model. To the best of our knowledge, this is the first work that considers both biopsy results and BI-RADS categories in mass malignancy classification.

3. Proposed Method

3.1. Datasets

Public dataset. We leverage the widely used DDSM (Digital Database for Screening Mammography) (Lee et al., 2017) as our public dataset. DDSM has 2,620 patient cases, each of which has standard four views of mammograms. Excluding some defective/corrupted cases, 2,578 cases (10,312 images in total) are used in this work. All cases are randomly divided at patient level into the training, validation, and test sets by approximately 8:1:1, resulting in 8,256, 1,020 and 1,036 images in the respective sets.

In-house dataset[2]. We collected and annotated mammographic data from Shenzhen People's Hospital in China to validate our proposed methods. The in-house dataset contains 2,749 patients' data taken with Siemens and Giotto equipment. After data cleaning, 2,807 four-view cases are obtained, consisting of normal, benign, and cancerous cases, which are close to the patient distribution in the hospital. All these mammograms are collected from digital radiography (DR) systems, which have better imaging quality than conventional computed radiography (CR) in the DDSM dataset. Lesion regions are first annotated by two radiologists and then reviewed by a senior reader. All cases are randomly split by 8:1:1 into the training, validation and test sets, each with 8,988, 1,120, and 1,120 images, respectively.

[2]This project is approved by the IRB number LL-XJS-2020011.

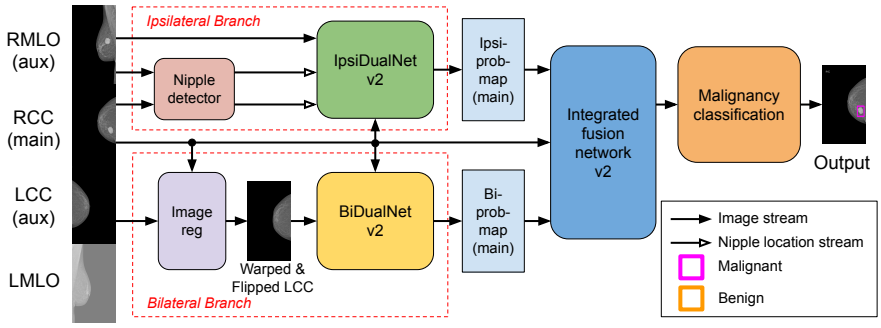


Figure 1: Framework of the proposed MommiNet-v2 for mass detection and malignancy classification.

3.2. *System Overview*

The framework of the our enhanced method, MommiNet-v2, is illustrated in Fig. 1. Different from our original MommiNet, a mass malignancy classification module is integrated. First, one input image is selected as the main view, and its corresponding ipsilateral and bilateral views are considered as the auxiliary views. These three images form the input of MommiNet-v2. As in Fig. 1, the main view ("RCC (main)") and the corresponding ipsilateral view ("RMLO (aux)") are input together into the ipsilateral branch. In parallel, the main view and the bilateral view ("LCC (aux)") are input into the bilateral branch. These two branches generate the probability maps of the main view, named as "Ipsi-prob map" and "Bi-prob map", respectively. Then, the probability maps along with the main view are fed into the integrated fusion network (v2) to generate the final mass detection results. Finally, the detected mass regions are further classified as benign or malignant by the malignancy classification module. More specifically, a DNN-based nipple detector is added to the ipsilateral branch to extract the nipple locations on both views ("RCC" and "RMLO") before inputting into *IpsiDualNet-v2*, and the bilateral view ("LCC") image is first registered towards the main view before input into the *BiDualNet-v2*. In practice, the proposed multi-view framework can be applied to any given view as the main image, and we apply it to all available views to obtain the mass detection and classification results on each view.

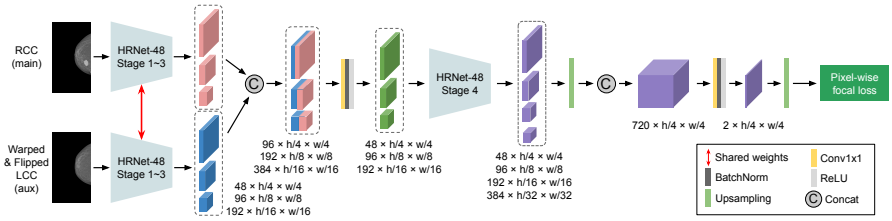


Figure 2: Architecture of BiDualNet-v2 for mass detection on bilateral dual-view mammograms.

3.3. Image Pre-Processing

Image Registration. To facilitate the DNN-based learning of the symmetry constraint from the bilateral images, we register the input pair of the same view images (e.g. two CC view images or two MLO view images). The auxiliary image is horizontally flipped and then warped toward the main image according to the breast contours. In particular, the nipple locations are used to roughly align the two MLO images before warping. A warped CC view example is shown in Fig. 1.

Nipple Detection. Nipple locations are required in image registration for MLO views and IpsiDualNet-v2. A Faster-RCNN based keypoint detector (Facebook, 2019) is trained to identify the nipple locations with satisfactory accuracy. For example, there is only one incorrect nipple prediction in our in-house dataset (11,228 images in total).

3.4. BiDualNet-v2

Most women have roughly symmetric breasts in terms of density and texture (Cunningham, 2013). This property is well leveraged by radiologists to identify the abnormalities in mammograms. Hinging on a bilateral dual-view, radiologists are able to locate a mass based on its distinct morphologic appearance and relative position compared to its corresponding area in the lateral image.

To incorporate this diagnostic prior information and facilitate the learning of the symmetry constraint, we develop *BiDualNet-v2* (*Bi*lateral *Dual*-view *Net*work-v2) as illustrated in Fig. 2. Compared with the original version of BiDualNet (Yang et al., 2020b), BiDualNet-v2 leverages the HRNet to learn the bilateral information. HRNet is capable of maintaining high-resolution image representations, which are not available in the previously adopted ResNet-based encoder structure. As HRNet's effectiveness

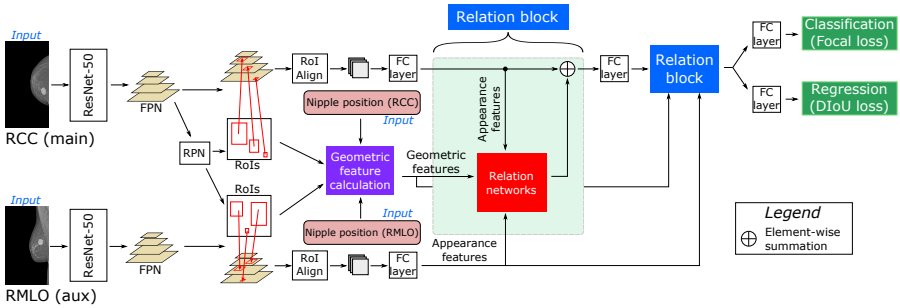


Figure 3: Architecture of IpsiDualNet-v2 for mass detection on ipsilateral mammograms.

has been proven in many common computer vision tasks, it has been already applied to medical image analysis (Xu et al., 2020; Huang et al., 2020). We utilize the HRNet structure in BiDualNet-v2 to better exploit the high-resolution symmetry information from the dual-view inputs. Our BiDualNet-v2 is derived from the HRNet-48 structure, enhanced with a Siamese input module and the pixel-wise focal loss (PWFL). The two Siamese inputs pass through the stage 1,2,3 simultaneously, and all these 3 stages share the same weights between the Siamese inputs, extracting features from the bilateral images in the same manner. The auxiliary feature map is then assumed as a reference and concatenated with the main feature map at the Stage 3, and in turn, the feature difference at the same location can highlight the abnormality. After a 1×1 convolution and the rest stages of HRNet, the segmentation network finally generates the feature map, which is converted into the probability map (Bi-prob-map) and input into the following integrated fusion network (v2). The model is optimized by minimizing the PWFL during training by performing focal loss (Lin et al., 2017) pixel by pixel between the ground-truth and the probability map, in which the focal loss itself tends to penalize more on those hard examples.

3.5. IpsiDualNet-v2

Ipsilateral images provide information on the same breast from two different views. Hence, a mass in the ipsilateral images tends to have similar distances to the nipple and share common appearance traits. This is an essential knowledge to assist radiologists in making decisions. We incorporate this prior diagnostic knowledge in our designed

IpsiDualNet-v2 (*Ipsi*lateral *Dual*-view *Net*work-v2) as presented in Fig. 3. Based on the Faster-RCNN detection architecture, we add the Siamese input module, a Feature Pyramid Networks (FPN) module (Facebook, 2019), and the designed relation blocks. The Siamese input module with FPN enables the two input branches to share the same weights and extract the features from the two ipsilateral views in the same way. In turn, the proposed relation blocks (described in the next paragraph) compute the appearance similarity and the geometry constraint between the RoIs from the two branches. Finally, the mass regions in the main image are detected and converted into a probability map. Moreover, focal loss (FL) (Lin et al., 2017) and Distance-IoU loss (DIoU) (Zheng et al., 2019) are used to improve the performance of *IpsiDualNet-v2*, and training with negative samples (normal cases) is enabled. Different from the normal IoU loss, the DIoU loss can better minimize the normalized distance between the target box and the anchor box. Compared with the original version IpsiDualNet (Yang et al., 2020b), the ResNet-50 backbone is replaced with the HRNet-48, which can better preserve the high resolution representations.

Relation networks Hu et al. (2017) explore the attention-based relationships (Vaswani et al., 2017) between two RoIs in single image based on the similarity of their appearance and geometric features, improving detection accuracy. Inspired by Hu et al. (2017), we develop a new relation block that enhances the appearance and geometric similarities of a lesion RoI in two ipsilateral images. The appearance similarity weight ω_A^{ij} between the i^{th} RoI $\mathbf{f}_m^i$ in the main image and the j^{th} RoI $\mathbf{f}_a^j$ in the auxiliary image is defined in Eq. (1), where two matrices W_a and W_m project the i^{th} and the j^{th} RoIs into subspaces to measure their appearance similarity. Regarding the geometric similarity, Eq. (2) considers RoIs' geometric factors $\mathbf{g}_t^k = \{d_t^k, w_t^k, h_t^k\}$, including the RoI-to-nipple distance, RoI width and height, where the subscript "t" indicates "m" or "a" (the main or auxiliary image). Other variables in Eqs. (1) and (2) have the same meaning as described in (Hu et al., 2017). The output of our relation block is also designed to add back to the main image stream without altering the feature dimension, and can be repeated for multiple times after the fully connected layer at the RoI head phase. We utilize two relation blocks as shown in Fig. 3 to emphasize the ipsilateral relationships. Radiologists routinely adopt the lesion-to-nipple distance as an important factor for es-

timating a lesion, since this distance is approximately the same in both CC and MLO views (Wei et al., 2009; Ikeda and Miyake, 2016). Fig. 4 shows an example of the similarity of the ROI-to-nipple distances.

$$\omega_A^{ij} = \frac{dot(W_a \mathbf{f}_a^j, W_m \mathbf{f}_m^i)}{\sqrt{D}}, \qquad i \in \{1, 2, ..., I\}, j \in \{1, 2, ..., J\}, \tag{1}$$

$$\mathcal{E}(\mathbf{g}_m^i, \mathbf{g}_a^j) = \mathcal{E}([\log(\frac{d_m^i}{d_a^j}), \log(\frac{w_m^i}{w_a^j}), \log(\frac{h_m^i}{h_a^j})]^T), \tag{2}$$

3.6. Integrated Fusion Network (v2)

We explore a fusion network to integrate the outputs of both ipsilateral and bilateral learning, and Fig. 5 illustrates the architecture of the designed integrated fusion network v2. The input of integrated fusion network consists of three images: the main image and the two probability maps generated by the IpsiDualNet-v2 and BiDualNet-v2 (shown in Fig. 1). These two probability maps are attentions of the comprehensive information from both bilateral and ipsilateral analysis. As shown in Fig. 5, there are two concatenation steps in the network. The first concatenation is to fuse the outputs from the two preceding sub-networks along with the main image. We perform the second concatenation in a U-net style, which can better keep the low level feature information along with the high level feature map generated by a series of convolutions, max-pooling and upsampling. Different from our original method in (Yang et al., 2020b), the backbone ResNet-50 is replaced with HRNet-48. Note that the HRNet-48 backbone comes from IpsiDualNet-v2 and is frozen during the training process. The final mass detection result is generated by this integrated fusion network.

3.7. Multi-Task Learning for Mass Malignancy Classification

While the MommiNet-v2 aims to detect mass regions in the high-resolution mammograms, a complete mass identification system is desirable to take one step further to predict the probability of the mass malignancy. This multi-stage framework design is consistent with radiologists' interpretation of a mammogram for diagnosing breast cancer.

Using biopsy results alone as labels that indicate normal/benign or malignant tissues already enables us to train a binary classifier to differentiate between benign and

Table 1: The definitions of BI-RADS and our pre-defined visual distances between neighboring BI-RADS categories. For example, the distance between BI-RADS 1 and 2 is set as 1.0, between 2 and 3 is set as 1.5, and so forth. The rightmost column shows the normalized values as a reference standard (or ground-truth (GT)) of the BI-RADS categories.

BI-RADS	Definition	Prob. of malignancy	Pre-defined distance	Normalized GT
0	Incomplete – need additional imaging evaluation	–	–	–
1	Normal	0%	1.0	0
2	Benign	0%	1.5	0.0926
3	Probably benign	<2%	2.0	0.2315
4	Suspicious for malignancy	4A: 2%-10%	1.5	0.4167
		4B:10%-50%	1.5	0.5556
		4C: 50%-95%	1.8	0.6944
5	Highly suggestive of malignancy	>95%	1.5	0.8611
6	Known biopsy-proven malignancy	100%	–	1

malignant mass patches. In addition, in most modern health care facilities, clinicians use BI-RADS to sort the assessment of breast lesions (Spak et al., 2017) into categories numbered 0 through 6. The brief definition of BI-RADS categories is introduced in Table 1, in which each BI-RADS category corresponds to a probability range of malignancy (American College of Radiology, 2013). In this regard, we investigate whether training with the combination of BI-RADS scores reported by radiologists and biopsy results could improve the automated classification performance of benign and malignant breast mass on mammographic images.

We propose a multi-task learning (Chen et al., 2019) module for benign and malignant mass classification on mammographic image patches, by training with both binary biopsy labels from pathology reports and BI-RADS scores from attending radiologists. The proposed module aims to improve the learning efficiency and prediction accuracy by learning two separate objectives from a shared representation. More specifically, a 121-layer densely connected convolutional network (Huang et al., 2017) is shared for feature learning, with a binary classification branch and a regression branch padded in parallel for benign/malignant classification and malignancy prediction, respectively.

During training, the BI-RADS scores range from 1 to 6 (4A, 4B, and 4C are treated as separate classes) are normalized into the range [0, 1] according to a pre-defined distance map. Since two neighboring BI-RADS categories are neither visually nor probabilistically equidistant, our collaborating radiologists approximately estimate the differences $l_{i,i+1}$ between any two neighboring BI-RADS categories i and $i+1$ and design a distance map upon them. The normalized distance between two neighboring categories $\bar{l}_{i,i+1}$ is defined as:

$$\bar{l}_{i,i+1} = \frac{l_{i,i+1}}{\sum_{i=1}^{7} l_{i,i+1}} \tag{3}$$

The pre-defined distance map and normalized value (as a reference standard ground-truth) of each BI-RADS category are shown in Table 1.

We design a multi-task learning loss combining binary classification and regression to simultaneously learn these two tasks in a unified architecture:

$$\mathcal{L}_{\text{MTL}} = \frac{1}{\sigma_1^2}\mathcal{L}_1(\mathbf{W}) + \frac{1}{2\sigma_2^2}\mathcal{L}_2(\mathbf{W}) + \log \sigma_1\sigma_2, \tag{4}$$

where $\mathcal{L}_1(\mathbf{W})$ is the binary cross entropy loss of the benign vs. malignant classification branch: $\mathcal{L}_1(\mathbf{W}) = -\log(\text{Softmax}(\mathbf{f}^{(\mathbf{W})}(\mathbf{x}), y_1))$, and $\mathcal{L}_2(\mathbf{W})$ is the Euclidean loss for the regression branch: $\mathcal{L}_2(\mathbf{W}) = \|\mathbf{f}^{\mathbf{W}}(\mathbf{x}) - y_2\|^2$. y_1 and y_2 are the ground-truth biopsy labels and the normalized BI-RADS scores, respectively. The multi-task learning module is optimized with respect to $\mathbf{W}$ as well as σ_1 and σ_2, where σ_1 and σ_2 are the standard deviation of the output values from the classification and regression branches, respectively. We follow Kendall et al. (Kendall et al., 2018) to optimize this objective function, which has been proven to be superior and more effective than manually tuning the relative weighting in a linear combination of each task's loss.

3.8. Training Strategies of the Framework

For the mass detection part, the detection subnetworks (i.e. the BiDualNet-v2 and the IpsiDualNet-v2) of MommiNet-v2 are trained separately. As shown in Fig. 1, we first train the IpsiDualNet-v2 and the BiDualNet-v2 sub-modules separately, and then convert their outputs into probability maps at the ipsilateral and bilateral branches,

respectively. Once the training procedure is completed, the two generated probability maps along with the main image stream are fed into the integrated fusion network (v2), which outputs the mass detection result. Since both the BiDualNet-v2 and the IpsiDualNet-v2 are trained independently, they function in parallel and generally do not influence each other. Nevertheless, the integrated fusion network (v2) is trained based on the preceding sub-networks' outputs, and therefore is subject to their performance.

The goal of mass malignancy classification in this work is to identify whether a detected region is malignant or not. Hence, we use both the gold patches annotated by radiologists (by referring to the biopsy results) and the patches detected by our MommiNet-v2 for training, but we only use the detected patches by our MommiNet-v2 model for validation and testing. To get the ground-truths of the detected patches, we match them with the gold patches annotated by radiologists. If the IoU (Intersection over Union) between a detected patch and a gold patch is equal or larger than 0.5, we set the ground-truth of this detected patch to associate with the gold patch. If the IoU between a detected patch and any gold patch is less than 0.5, we set the ground-truth of this detected patch as not malignant (normal or benign). All image patches are resized to 448 $\times$ 448 pixels. We augment the training data using random horizontal and vertical flipping, random brightness and contrast adjustment, and random affine transformation. We initialize the DenseNet-121 (Huang et al., 2017) with the network weights pre-trained on the ImageNet classification task. We set the batch size to 16 and the initial learning rate as 0.001 and reduce it by a factor of 0.1 after the loss plateaued for 5 epochs. Stochastic gradient descent (SGD) optimizer with a momentum of 0.9 is used to optimize the training. Early stopping with a maximum running of 100 training epochs is used to avoid overfitting.

Although we processed our in-house dataset to ensure each case has four views, our framework can still work with missing views, as the system automatically degrades into the corresponding dual-view model.

Table 2: Ablation study on ipsilateral and bilateral branches the DDSM dataset. 95% confidence intervals (CI) are shown in the square brackets.

Networks	**Ipsilateral** (Recall@FPPI)			**Bilateral** (Recall@FPPI)		
	R@0.5	**R@1.0**	**R@2.0**	**R@0.5**	**R@1.0**	**R@2.0**
BiDualNet	0.668 [0.664, 0.672]	0.809 [0.803, 0.815]	0.887 [0.883, 0.891]	0.783 [0.779, 0.787]	0.842 [0.838, 0.846]	0.891 [0.887, 0.895]
BiDualNet-v2	0.704 [0.701, 0.707]	0.803 [0.800, 0.806]	0.884 [0.879, 0.889]	**0.801** [0.799, 0.803]	**0.853** [0.851, 0.855]	**0.891** [0.888, 0.894]
IpsiDualNet w/o Relation Blocks	0.679 [0.675, 0.683]	0.772 [0.768, 0.776]	0.838 [0.832, 0.844]	0.734 [0.728, 0.740]	0.786 [0.781, 0.791]	0.835 [0.830, 0.840]
IpsiDualNet	0.764 [0.762, 0.766]	0.828 [0.824, 0.832]	0.879 [0.875, 0.883]	0.652 [0.646, 0.658]	0.747 [0.741, 0.753]	0.824 [0.818, 0.830]
IpsiDualNet-v2	**0.811** [0.805, 0.817]	**0.843** [0.837, 0.849]	**0.889** [0.885, 0.893]	0.678 [0.674, 0.682]	0.764 [0.760, 0.768]	0.801 [0.795, 0.807]

Table 3: Ablation study on ipsilateral and bilateral branches on our in-house dataset. 95% confidence intervals (CI) are shown in the square brackets.

Networks	**Ipsilateral** (Recall@FPPI)			**Bilateral** (Recall@FPPI)		
	R@0.5	**R@1.0**	**R@2.0**	**R@0.5**	**R@1.0**	**R@2.0**
BiDualNet	0.709 [0.705, 0.713]	0.782 [0.778, 0.786]	0.898 [0.892, 0.904]	0.874 [0.870, 0.878]	0.931 [0.927, 0.935]	0.948 [0.944, 0.952]
BiDualNet-v2	0.741 [0.738, 0.746]	0.802 [0.799, 0.807]	0.884 [0.880, 0.888]	**0.892** [0.889, 0.895]	**0.932** [0.930, 0.934]	**0.950** [0.947, 0.953]
IpsiDualNet w/o Relation Blocks	0.804 [0.801, 0.807]	0.856 [0.853, 0.859]	0.908 [0.903, 0.913]	0.828 [0.823, 0.833]	0.881 [0.876, 0.886]	0.917 [0.913, 0.921]
IpsiDualNet	0.882 [0.878, 0.886]	0.917 [0.913, 0.921]	0.958 [0.953, 0.963]	0.777 [0.771, 0.783]	0.832 [0.826, 0.838]	0.903 [0.826, 0.838]
IpsiDualNet-v2	**0.891** [0.888, 0.894]	**0.933** [0.930, 0.936]	**0.961** [0.957, 0.965]	0.788 [0.785, 0.791]	0.854 [0.851, 0.857]	0.897 [0.894, 0.900]

4. Experiments

In this section, we perform extensive experiments on the DDSM and our in-house datasets to validate our proposed method. The model on each dataset is independently trained based on the pre-trained ImageNet model (Krizhevsky et al., 2012). The recall at different numbers of false positive per image (FPPI), namely, the free-response receiver operating characteristic (FROC) is selected as our evaluation metric to compare with the previous work. Every image is resized to at most 3000×1500 according to the aspect ratio as input. A mass is assumed as successfully identified if the IoU of the predicted output and the ground truth mask is greater than 0.2, as commonly used in

Table 4: Impact of different geometric features on prediction performance on our in-house dataset.

Geometric Features	**R@0.5**	**R@1.0**	**R@2.0**
Shape and location of RoI (i.e., Eq. (2) in (Ma et al., 2019))	0.86	0.90	0.93
Dummy nipple point (Central point of every image)	0.80	0.85	0.89
RoI-to-nipple distance (Ours, in IpsiDualNet-v2)	**0.891**	**0.933**	**0.961**

Table 5: Comparison of concatenation location of HRNet-48 in BiDualNet-v2 on the DDSM and in-house datasets (referring to Fig. 2).

Datasets	**Concatenation Stage of HRNet-48**	**R@0.5**	**R@1.0**	**R@2.0**
DDSM	Concat @ Stage 1	0.763	0.831	0.874
	Concat @ Stage 2	0.790	0.852	0.878
	Concat @ Stage 3 (Default in BiDualNet-v2)	**0.801**	**0.853**	**0.891**
In-house	Concat @ Stage 1	0.850	0.912	0.918
	Concat @ Stage 2	0.871	0.917	0.940
	Concat @ Stage 3 (Default in BiDualNet-v2)	**0.892**	**0.932**	**0.950**

earlier studies (Agarwal et al., 2019; Dhungel et al., 2017).

4.1. Results of Mass Detection

4.1.1. Ablation Study

Ipsilateral and Bilateral Learning. We train and test BiDualNet (and v2), IpsiDualNet (and v2) and their degraded versions "IpsiDualNet w/o Relation Blocks" (the two feature streams from the main and auxiliary images are directly concatenated after the RoI alignment stage) on both ipsilateral and bilateral images.

Table 2 shows the results on DDSM. It can be observed that BiDualNet always achieves the relatively higher recall scores on bilateral images, and IpsiDualNet generally has better performance on ipsilateral images. The results of BiDualNet-v2 and IpsiDualNet-v2 surpass their original versions at the appropriate lateral sides. We believe this is due to the fact that the high-resolution features of smaller mass lesions are

Table 6: Performance comparison of various methods on the DDSM dataset. *CVR-RCNN*: Cross-View Relation Region-based Convolutional Neural Network; *CBN*: Contrasted Bilateral Network 95% confidence intervals (CI) are shown in the square brackets.

View	Method	DDSM (train/val/test)	Recall@FPPI R@0.5	R@1.0	R@2.0
Single	Campanini et al. (Campanini et al., 2004)	1400/_/512	~0.54	~0.74	~0.86
	Nazaré Silva et al. (de Nazaré Silva et al., 2015)	349/150/100	n/a	~0.8033	n/a
	Faster-RCNN (Liu et al., 2019)	80%/10%/10%	0.6610	0.7246	0.7839
	Mask-RCNN (Liu et al., 2019)		0.6441	0.7458	0.8178
	DeepLab+NL+PWFL	8256/1020/1036	0.68	0.78	0.83
	HRNet+PWFL		0.70	0.79	0.84
	Faster-RCNN(ResNet-50)+FPN+FL+DIoU		0.74	0.82	0.88
	Faster-RCNN(HRNet-48)+FPN+FL+DIoU		0.76	0.82	0.88
Dual	CVR-RCNN (Ma et al., 2019)	410/_/102	n/a	n/a	~0.88
	CBN (Liu et al., 2019)	80%/10%/10%	0.6907	0.7881	0.8559
	BG-RCNN (Liu et al., 2020)	1638/205/205	0.795	0.866	0.918
	BiDualNet	8256/1020/1036	0.783	0.842	0.891
	BiDualNet-v2		0.801	0.853	0.891
	IpsiDualNet		0.764	0.828	0.879
	IpsiDualNet-v2		0.811	0.843	0.889
Tri	MommiNet	8256/1020/1036	0.802 [0.799, 0.805]	0.849 [0.846, 0.852]	0.892 [0.890, 0.894]
	MommiNet-v2	8256/1020/1036	0.831 [0.827, 0.835]	0.850 [0.848, 0.852]	0.898 [0.894, 0.902]

better captured by the HRNet structure. It is also clear that IpsiDualNet outperforms IpsiDualNet w/o Relation Blocks on ipsilateral images, which suggests that the designed relation module remarkably enhances IpsiDualNet. Thus, BiDualNet and IpsiDualNet are respectively applied to the bilateral and ipsilateral analysis in MommiNet. The results on our in-house dataset in Table 3 follow a similar trend as on DDSM. The efficacy of IpsiDualNet and BiDualNet on the respective ipsilateral and bilateral sides is even more substantial. Similar to Table 2, BiDualNet-v2 and IpsiDualNet-v2 outperform the original versions at the proposed lateral sides.

Geometric Features in IpsiDualNet-v2. Table 4 shows the impact of different geometric features on IpsiDualNet-v2, including the features in (Ma et al., 2019), the dummy nipple, and our RoI-to-nipple-distance based features. It clearly demonstrates that the RoI-to-nipple-distance based geometric features generate the best performance

of IpsiDualNet-v2.

Concatenation Stage in BiDualNet-v2. In our proposed BiDualNet-v2 (shown in Fig. 2), we concatenate the bilateral feature maps at Stage 3 of HRNet-48. This stage is also the last concatenation position in BiDualNet-v2 due to our 48-GB GPU memory limitation, since any later concatenation stage requires the GPU memory beyond the limit. We here investigate the impact of different concatenation stages in BiDualNet-v2. Table 5 lists the results on both the entire DDSM and our in-house datasets when the dual feature maps are concatenated at Stage 1, 2, or 3 of HRNet-48, respectively. The results show that our default concatenation operation at Stage 3 in BiDualNet-v2 achieves the best result on both datasets.

4.1.2. Results on the DDSM and In-House Datasets

Table 6 compares the performance of various single-view, dual-view and tri-view methods on the DDSM dataset. The approaches in the references (Campanini et al., 2004; de Nazaré Silva et al., 2015; Liu et al., 2019; Ma et al., 2019), which reported evaluation on DDSM with normal patients data using the FROC metric, are selected as the comparison methods. Among the single-view methods, our Faster-RCNN with the HRNet-48 backbone and the modules of FPN, FL, and DIoU loss achieves the best result. Regarding the recent dual-view methods, even though BG-RCNN (Liu et al., 2020) provides higher recall results at FPPI=1.0/2.0, our IpsiDualNet-v2 and tri-view models achieve better recall rates at FPPI=0.5. In addition, only a small subset of DDSM (512 cases) is used in the BG-RCNN study, while the entire DDSM (2578 cases excluding the corrupt files) is included in our study. Finally, our proposed tri-view MommiNet-v2 surpass all our dual-view baselines and improve the performance of the original MommiNet. To the best of our knowledge, MommiNet-v2 achieves the highest recall scores on the entire DDSM dataset.

Fig. 6 shows an example case of mass detection on the public DDSM dataset using a single-view method, MommiNet and MommiNet-v2. For the single view method, a mass is missed in the LMLO image, and a false positive is predicted in the LCC image. MommiNet eliminates the false positive in the LCC image but still misses the mass in the LMLO image. In comparison, MommiNet-v2 successfully predicts all masses

Table 7: Performance comparison of various methods on the in-house dataset. 95% confidence intervals (CI) are shown in the square brackets.

View Type	Method	R@0.5	R@1.0	R@2.0
Single-View	DeepLab+NL+PWFL	0.81	0.84	0.90
	HRNet+PWFL	0.83	0.87	0.90
	Faster-RCNN(ResNet-50)+FPN+FL+DIoU	0.82	0.89	0.91
	Faster-RCNN(HRNet-48)+FPN+FL+DIoU	0.84	0.90	0.91
Dual-View	BiDualNet	0.874	0.931	0.948
	BiDualNet-v2	0.892	0.932	0.950
	IpsiDualNet	0.882	0.917	0.958
	IpsiDualNet-v2	0.891	0.933	0.961
Tri-View	MommiNet	0.901 [0.897, 0.905]	0.939 [0.935, 0.943]	0.960 [0.957, 0.963]
	MommiNet-v2	**0.912** [0.908, 0.916]	**0.939** [0.936, 0.942]	**0.962** [0.959, 0.965]

without any false positives.

Various methods are also tested on the in-house dataset, as shown in Table 7. The dual-view networks are constantly better than the single-view methods, and the tri-view MommiNet-v2 again achieves the best result at all FPPIs. Furthermore, due to the DR images' better quality, the results on the in-house dataset are generally better than on DDSM, and the proposed method achieves remarkably higher recall rates on the in-house dataset.

4.2. Results of Mass Malignancy Classification

We train and evaluate our proposed mass malignancy classification module using multi-task learning on the in-house dataset. The proposed module is trained and cross-validated on a combination of 1,173 mammographic image patches detected by our MommiNet-v2 and the gold patches annotated by the radiologists. A hold-out set containing 390 image patches detected by the MommiNet-v2 (no patient overlap with the training and validation sets) is used to evaluate the performance of the proposed multi-task learning method.

Our multi-task learning method achieves an AUC of 0.9144 (95% CI [0.9112, 0.9175]) for benign versus malignant mass classification, compared to 0.8860 (95% CI

[0.8816, 0.8908]) using only binary labels in training ($p<0.05$). We show the receiver operating characteristic (ROC) curves of two random testing runs in Fig. 7. This proves that in addition to the objective biopsy labels, subjective BI-RADS scores can provide auxiliary information in training deep neural networks for benign/malignant mass classification or malignancy prediction in a multi-task setting. Furthermore, we compare the effectiveness of manual weight tuning and automatic weight learning (shown in Fig. 8). The optimal performance for manual weight tuning (i.e., AUC=0.8981) is achieved when the biopsy label-based classification branch accounts for 70%, and the BI-RADS score-based regression branch accounts for 30% of the total weight, respectively. The automatic weight learning strategy achieves better overall performance (i.e., AUC=0.9144) and avoids expensive, time-consuming manual selection of individual weight.

4.3. Visualization Results

Fig. 9 shows an example case of mass detection on our in-house dataset using single-view method, MommiNet and MommiNet-v2. For the single view method, a mass is missed in the LMLO image, and a false positive is predicted in the LCC image. MommiNet detects the mass in the LMLO image but still has the false positive in the LCC image. In comparison, MommiNet-v2 successfully predicts all masses with the highest IoU values, and without any false positive. Furthermore, MommiNet-v2 correctly classifies the predicted masses as malignant ones.

5. Discussion

We take a multi-stage approach toward a CAD system for mammograms, and MommiNet-v2 is an integral part of the system, targeting for mass detection and malignancy classification. Our mammographic breast lesion diagnostic system has already been deployed in the collaborating hospital, and the initial feedback from the radiologists has been encouraging in terms of the overall accuracy of mass detection and malignancy classification (Yang et al., 2020a).

However, there are some limitations in this study. Firstly, for very rare cases, nipples can not be clearly observed, leading to the inaccurate nipple location estimation which may compromise the performance of the ipsiDualNet-v2 branch.

Secondly, our current in-house dataset consists of data generated by equipment from two different vendors, and are collected from one medical institute. The training and testing data are predominantly labeled by our collaborating radiologists. Therefore our data and results are subject to the patient distribution and radiologists expertise from our collaborating hospital. Moving forward, effort has been undertaken to construct a larger scale multi-center dataset.

Lastly, the malignancy classification stage is influenced by the detection results of the MommiNet-v2 since the input of the classification task is the output of the detection. If a mass is missed by MommiNet-v2, it will not be classified in the malignancy classification stage. One possible way to remedy this is to use a sliding window to classify all the patches. But we did not perform this since the goal in this study is to determine the malignancy of the detected patches in the first stage. In addition, for mass malignancy classification, we only utilize image appearances as visual features to train deep learning models, while linked electronic health records contain richer clinical information such as age, breast radiology history, family history and symptoms, and could help improve cancer prediction accuracy, as in Akselrod-Ballin et al. (Akselrod-Ballin et al., 2019). Our mass malignancy classification model could also benefit from this approach.

6. Conclusion and Future Work

In this paper, we further enhance the first multi-view DNN architecture MommiNet into MommiNet-v2 to perform joint ipsilateral and bilateral analysis on mammograms for high precision mass detection and malignancy classification. By carefully designing the DNN architecture, MommiNet-v2 can effectively learn the geometry constraint and symmetry constraint from the ipsilateral and bilateral views respectively. Its efficacy can be further verified by our extensive experiment results and the SOTA FROC performance achieved on both the DDSM dataset and our in-house dataset. The proposed

multi-task learning strategy has also shown great potential for mass malignancy classification. We plan to further improve the system performance with additional modality data, such as patients' health records, ultrasound and MRI etc. We are also expanding our in-house datasets to include data from more medical providers.

Declaration of Competing Interest

The authors have no competing interest to declare.

CRediT Authorship Contribution Statement

Zhicheng Yang: Conceptualization, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing - original draft, review & editing. **Zhenjie Cao:** Conceptualization, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing - original draft, review & editing. **Yanbo Zhang:** Conceptualization, Data Curation, Investigation, Methodology, Software, Writing - original draft, review & editing. **Yuxing Tang:** Conceptualization, Formal analysis, Investigation, Methodology, Software, Visualization, Writing - original draft, review & editing. **Xiaohui Lin:** Investigation, Data Curation, Validation, Writing - review. **Rushan Ouyang:** Investigation, Data Curation, Validation, Writing - review. **Mingxiang Wu:** Investigation, Data Curation, Validation, Writing - review. **Mei Han:** Conceptualization, Investigation, Resources, Writing - review, Supervision. **Jing Xiao:** Investigation, Resources, Writing - review, Supervision. **Lingyun Huang:** Resources, Writing - review. **Shibin Wu:** Methodology, Writing - review. **Jie Ma:** Conceptualization, Investigation, Data Curation, Resources, Writing - review, Supervision. **Peng Chang:** Conceptualization, Investigation, Methodology, Resources, Writing - original draft, review & editing, Supervision.

Acknowledgments

This research has been supported by the Shenzhen Science and Technology Research Fund No. JCYJ20180305164740612.

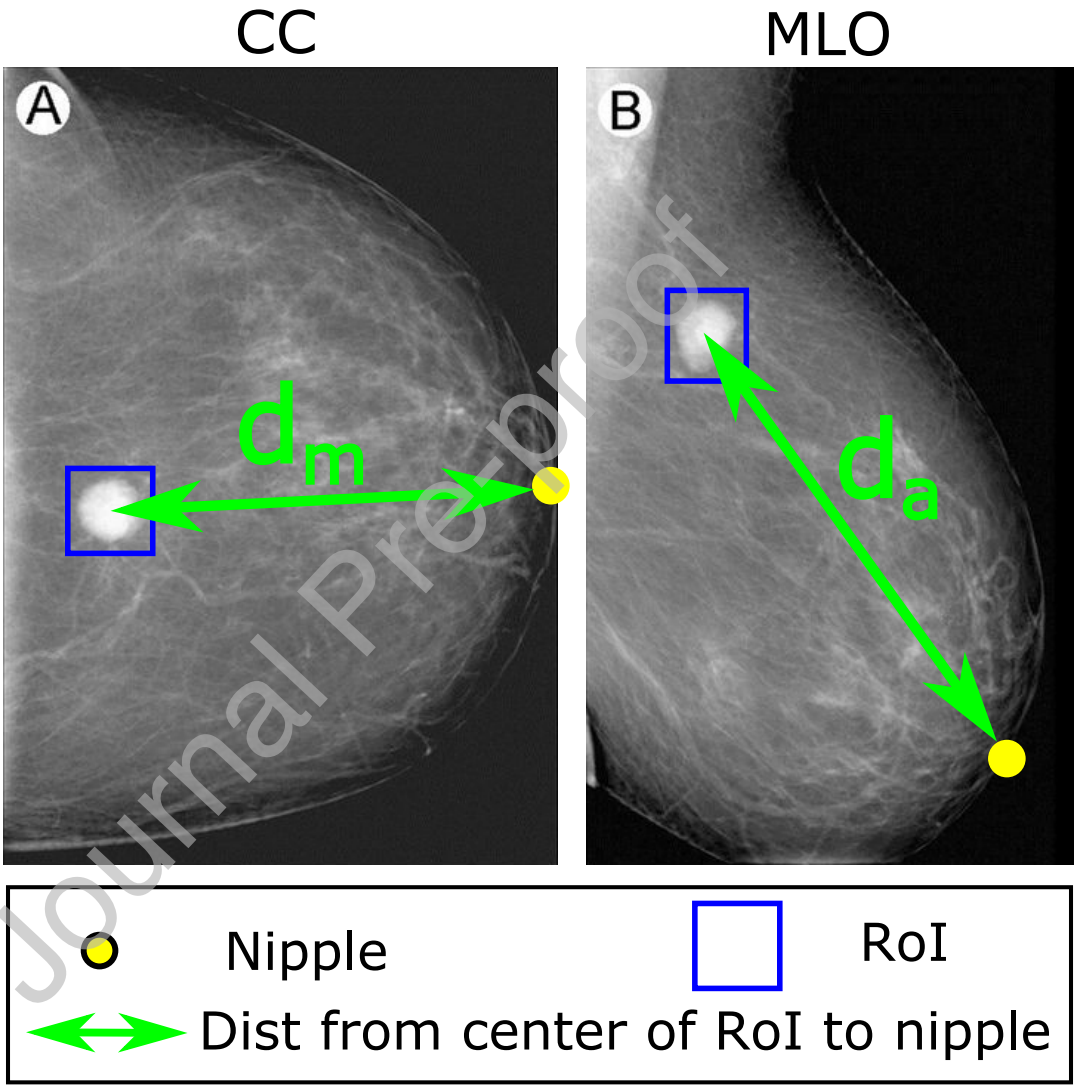


Figure 4: Similarity of RoI-to-nipple distances.

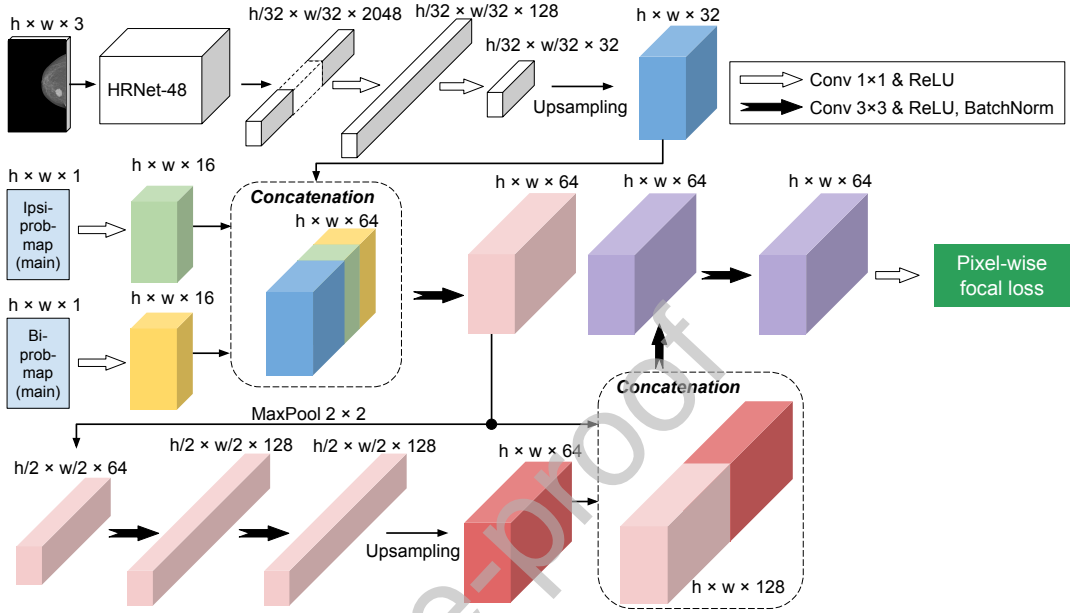


Figure 5: Integrated fusion network (v2).

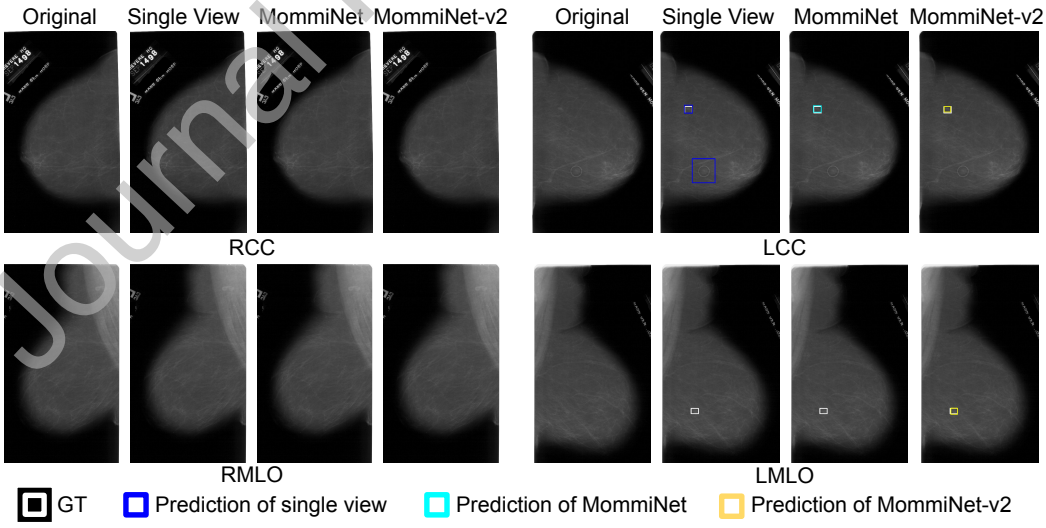


Figure 6: An example case of mass detection on the DDSM dataset using a single view method (Faster-RCNN) and our MommiNet and MommiNet-v2 on DDSM. GTs are shown in white bounding boxes.

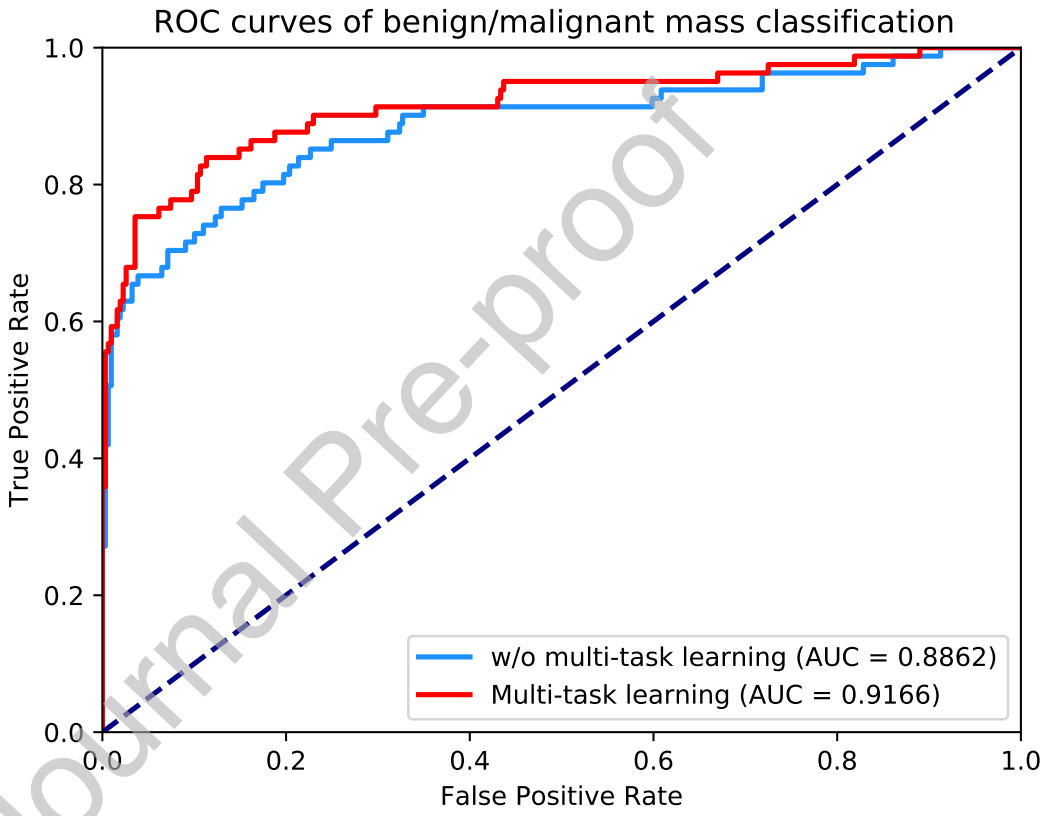


Figure 7: ROC curves of mass malignancy classification using both biopsy labels and BI-RADS scores (multi-task learning) compared to using only biopsy labels (without multi-task learning).

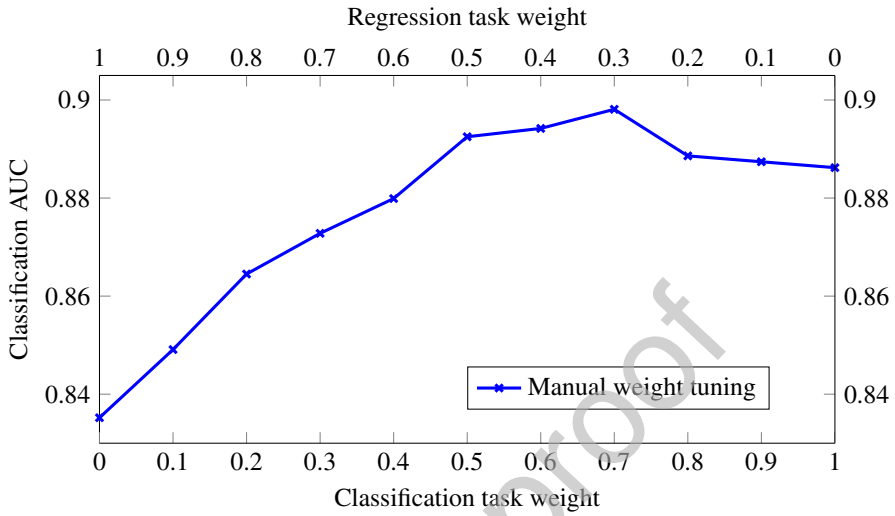


Figure 8: Mass malignancy classification performance comparison between manual weight tuning and automatic weight learning for multi-task learning. For manual weight tuning, the sum of classification and regression task weight is 1. We plot different weight combinations in the blue color. The maximal AUC using manual weight tuning is 0.8981. The automatic weight learning method has nothing to do with individual weight, which achieves a mean AUC of 0.9144.

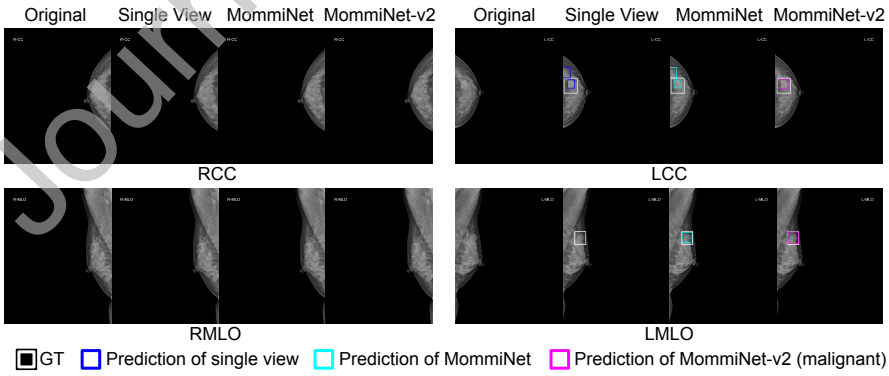


Figure 9: An example case of mass detection using a single view method (Faster-RCNN) and our MommiNet and MommiNet-v2 on the in-house dataset. GTs are shown in white bounding boxes.

References

Abdelhafiz, D., Nabavi, S., Ammar, R., Yang, C., Bi, J., 2019. Residual Deep Learning System for Mass Segmentation and Classification in Mammography, in: Proceedings of the 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics, pp. 475–484.

Agarwal, R., Diaz, O., Llad, X., Yap, M.H., Mart, R., 2019. Automatic mass detection in mammograms using deep convolutional neural networks. Journal of Medical Imaging 6, 1 – 9.

Akselrod-Ballin, A., Chorev, M., Shoshan, Y., Spiro, A., Hazan, A., Melamed, R., Barkan, E., Herzel, E., Naor, S., Karavani, E., et al., 2019. Predicting breast cancer by applying deep learning to linked health records and mammograms. Radiology 292, 331–342.

Al-Antari, M.A., Al-Masni, M.A., Choi, M.T., Han, S.M., Kim, T.S., 2018. A fully integrated computer-aided diagnosis system for digital X-ray mammograms via deep learning detection, segmentation, and classification. International journal of medical informatics 117, 44–54.

Al-Masni, M.A., Al-Antari, M.A., Park, J.M., Gi, G., Kim, T.Y., Rivera, P., Valarezo, E., Choi, M.T., Han, S.M., Kim, T.S., 2018. Simultaneous detection and classification of breast masses in digital mammograms via a deep learning YOLO-based CAD system. Computer methods and programs in biomedicine 157, 85–94.

American College of Radiology, 2013. ACR BI-RADS Atlas: Breast Imaging Reporting and Data System; Mammography, Ultrasound, Magnetic Resonance Imaging, Follow-up and Outcome Monitoring, Data Dictionary. ACR, American College of Radiology.

Campanini, R., Dongiovanni, D., Iampieri, E., Lanconelli, N., Masotti, M., Palermo, G., Riccardi, A., Roffilli, M., 2004. A novel featureless approach to mass detection in digital mammograms based on support vector machines. Physics in Medicine & Biology 49, 961.

Cao, Z., Yang, Z., Zhang, Y., Lin, R.S., Wu, S., Huang, L., Han, M., Ma, J., 2019a. Deep learning based mass detection in mammograms., in: GlobalSIP, pp. 1–5.

Cao, Z., Yang, Z., Zhuo, X., Lin, R.S., Wu, S., Huang, L., Han, M., Zhang, Y., Ma, J., 2019b. Deeplima: Deep learning based lesion identification in mammograms, in: Proceedings of the IEEE International Conference on Computer Vision Workshops.

Carneiro, G., Nascimento, J., Bradley, A.P., 2017. Automated analysis of unregistered multi-view mammograms with deep learning. IEEE transactions on medical imaging 36, 2355–2365.

Caruana, R., 1997. Multitask learning. Machine learning 28, 41–75.

Chen, Q., Peng, Y., Keenan, T., Dharssi, S., Agro, E., et al., 2019. A multi-task deep learning model for the classification of age-related macular degeneration. AMIA Summits on Translational Science Proceedings 2019, 505.

Cunningham, D., 2013. The Ups and Downs of Breasts, Physicians & Midwives. https://physiciansandmidwives.com/2013/12/11/ups-and-downs-of-breasts/.

Dhungel, N., Carneiro, G., Bradley, A.P., 2017. A deep learning approach for the analysis of masses in mammograms with minimal user intervention. Medical image analysis 37, 114–128.

Diniz, J.O.B., Diniz, P.H.B., Valente, T.L.A., Silva, A.C., de Paiva, A.C., Gattass, M., 2018. Detection of mass regions in mammograms by bilateral analysis adapted to breast density using similarity indexes and convolutional neural networks. Computer methods and programs in biomedicine 156, 191–207.

Facebook, 2019. Fast, modular reference implementation of Instance Segmentation and Object Detection algorithms in PyTorch. https://github.com/facebookresearch/maskrcnn-benchmark.

Guan, Q., Huang, Y., Zhong, Z., Zheng, Z., Zheng, L., Yang, Y., 2020. Thorax disease classification with attention guided convolutional neural network. Pattern Recognition Letters 131, 38–45.

Hu, H., Gu, J., Zhang, Z., Dai, J., Wei, Y., 2017. Relation networks for object detection. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition , 3588–3597.

Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q., 2017. Densely connected convolutional networks, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4700–4708.

Huang, Z., Lin, J., Xu, L., Wang, H., Bai, T., Pang, Y., Meen, T.H., 2020. Fusion high-resolution network for diagnosing chestx-ray images. Electronics 9, 190.

Ikeda, D., Miyake, K.K., 2016. Breast imaging: the requisites E-book. Elsevier Health Sciences.

Jiao, Z., Gao, X., Wang, Y., Li, J., 2016. A deep feature based framework for breast masses classification. Neurocomputing 197, 221–231.

Kendall, A., Gal, Y., Cipolla, R., 2018. 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, pp. 7482–7491.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks, in: Advances in neural information processing systems, pp. 1097–1105.

Lee, R.S., Gimenez, F., Hoogi, A., Miyake, K.K., Gorovoy, M., Rubin, D.L., 2017. A curated mammography data set for use in computer-aided detection and diagnosis research. Scientific data 4, 170177.

Li, H., Chen, D., Nailon, W.H., Davies, M.E., Laurenson, D., 2019. A deep dual-path network for improved mammogram image processing, in: ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE. pp. 1224–1228.

Li, Y., Chen, H., Zhang, L., Cheng, L., 2018. Mammographic mass detection based on convolution neural network, in: 2018 24th International Conference on Pattern Recognition (ICPR), IEEE. pp. 3850–3855.

Li, Y., Zhang, L., Chen, H., Cheng, L., 2020. Mass detection in mammograms by bilateral analysis using Convolution Neural Network. Computer Methods and Programs in Biomedicine , 105518.

Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P., 2017. Focal loss for dense object detection, in: Proceedings of the IEEE international conference on computer vision, pp. 2980–2988.

Liu, Y., Zhang, F., Zhang, Q., Wang, S., Wang, Y., Yu, Y., 2020. Cross-View Correspondence Reasoning Based on Bipartite Graph Convolutional Network for Mammogram Mass Detection, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 3812–3822.

Liu, Y., Zhou, Z., Zhang, S., Luo, L., Zhang, Q., Zhang, F., Li, X., Wang, Y., Yu, Y., 2019. From unilateral to bilateral learning: Detecting mammogram masses with contrasted bilateral network, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer. pp. 477–485.

Ma, J., Liang, S., Li, X., Li, H., Menze, B.H., Zhang, R., Zheng, W.S., 2019. Cross-view relation networks for mammogram mass detection. ArXiv abs/1907.00528.

McKinney, S., Sieniek, M., Godbole, V., Godwin, J., Antropova, N., Ashrafian, H., Back, T., Chesus, M., Corrado, G., Darzi, A., Etemadi, M., Garcia-Vicente, F., Gilbert, F., Halling-Brown, M., Hassabis, D., Jansen, S., Karthikesalingam, A., Kelly, C., King, D., Shetty, S., 2020. International evaluation of an ai system for breast cancer screening. Nature 577, 89–94.

de Nazaré Silva, J., de Carvalho Filho, A.O., Silva, A.C., De Paiva, A.C., Gattass, M., 2015. Automatic detection of masses in mammograms using quality threshold clustering, correlogram function, and svm. Journal of digital imaging 28, 323–337.

Pedro, R.W.D., Machado-Lima, A., Nunes, F.L., 2019. Is mass classification in mammograms a solved problem?-a critical review over the last 20 years. Expert Systems with Applications 119, 90–103. Publisher: Elsevier.

Perek, S., Hazan, A., Barkan, E., Akselrod-Ballin, A., 2018. Siamese network for dual-view mammography mass matching, in: Image Analysis for Moving Organ, Breast, and Thoracic Images. Springer, pp. 55–63.

Rangayyan, R.M., El-Faramawy, N.M., Desautels, J.L., Alim, O.A., 1997. Measures of acutance and shape for classification of breast tumors. IEEE Transactions on medical imaging 16, 799–810.

Ren, S., He, K., Girshick, R., Sun, J., 2015. Faster r-cnn: Towards real-time object detection with region proposal networks, in: Advances in neural information processing systems, pp. 91–99.

Ren, Y., Hou, R., Kong, D., Geng, Y., Grimm, L.J., Marks, J.R., Lo, J.Y., 2019. Multiview mammographic mass detection based on a single shot detection system, in: Medical Imaging 2019: Computer-Aided Diagnosis, International Society for Optics and Photonics. p. 109500E.

Sahiner, B., Chan, H.P., Hadjiiski, L.M., Helvie, M.A., Paramagul, C., Ge, J., Wei, J., Zhou, C., 2006. Joint two-view information for computerized detection of microcalcifications on mammograms. Medical physics 33, 2574–2585.

Spak, D.A., Plaxco, J., Santiago, L., Dryden, M.J., Dogan, B., 2017. Bi-rads® fifth edition: A summary of changes. Diagnostic and interventional imaging 98, 179–190.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I., 2017. Attention is all you need, in: Advances in neural information processing systems, pp. 5998–6008.

Wang, C.R., Li, J., Zhang, F., Sun, X., Dong, H., Yu, Y., Wang, Y., 2020a. Bilateral asymmetry guided counterfactual generating network for mammogram classification. arXiv preprint arXiv:2009.14406 .

Wang, C.R., Zhang, F., Yu, Y., Wang, Y., 2020b. Br-gan: Bilateral residual generating adversarial network for mammogram classification, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer. pp. 657–666.

Wang, D., Shi, L., Heng, P.A., 2009. Automatic detection of breast cancers in mammograms using structured support vector machines. Neurocomputing 72, 3296–3302.

Wang, H., Feng, J., Zhang, Z., Su, H., Cui, L., He, H., Liu, L., 2018. Breast mass classification via deeply integrating the contextual information from multi-view data. Pattern Recognition 80, 42 – 52.

Wang, J., Sun, K., Cheng, T., Jiang, B., Deng, C., Zhao, Y., Liu, D., Mu, Y., Tan, M., Wang, X., et al., 2020c. Deep high-resolution representation learning for visual recognition. IEEE transactions on pattern analysis and machine intelligence .

Wei, J., Chan, H.P., Sahiner, B., Zhou, C., Hadjiiski, L.M., Roubidoux, M.A., Helvie, M.A., 2009. Computer-aided detection of breast masses on mammograms: Dual system approach with two-view analysis. Medical physics 36, 4451–4460.

Wu, N., Phang, J., Park, J., Shen, Y., Huang, Z., Zorin, M., Jastrzebski, S., Fevry, T., Katsnelson, J., Kim, E., Wolfson, S., Parikh, U., Gaddam, S., Lin, L., Ho, K., Weinstein, J., Reig, B., Gao, Y., Pysarenko, H., Geras, K., 2019. Deep neural networks improve radiologists performance in breast cancer screening. IEEE Transactions on Medical Imaging PP, 1–1.

Xi, P., Shu, C., Goubran, R., 2018. Abnormality detection in mammography using deep convolutional neural networks, in: 2018 IEEE International Symposium on Medical Measurements and Applications (MeMeA), IEEE. pp. 1–6.

Xu, S., Lu, H., Ye, M., Yan, K., Zhu, W., Jin, Q., 2020. Improved cascade r-cnn for medical images of pulmonary nodules detection combining dilated hrnet, in: Proceedings of the 2020 12th International Conference on Machine Learning and Computing, pp. 283–288.

Yang, Z., Cao, Z., Zhang, Y., Chang, P., Wu, S., Huang, L., Xu, W., Han, M., Xiao, J., Ma, J., 2020a. Mabel: An ai-powered mammographic breast lesion diagnostic system, in: 2020 IEEE International Conference on E-health Networking, Application & Services (Healthcom).

Yang, Z., Cao, Z., Zhang, Y., Han, M., Xiao, J., Huang, L., Wu, S., Ma, J., Chang, P., 2020b. Momminet: Mammographic multi-view mass identification networks, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer. pp. 200–210.

Zhang, F., Luo, L., Sun, X., Zhou, Z., Li, X., Yu, Y., Wang, Y., 2019. Cascaded generative and discriminative learning for microcalcification detection in breast mammograms, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 12578–12586.

Zheng, Z., Wang, P., Liu, W., Li, J., Ye, R., Ren, D., 2019. Distance-iou loss: Faster and better learning for bounding box regression. arXiv preprint arXiv:1911.08287 .

Zhu, J.Y., Park, T., Isola, P., Efros, A.A., 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks, in: Proceedings of the IEEE International Conference on Computer Vision, pp. 2223–2232.

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT Authorship Statement

Zhicheng Yang:
Conceptualization, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing - original draft, review & editing.

Zhenjie Cao:
Conceptualization, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing - original draft, review & editing.

Yanbo Zhang:
Conceptualization, Data Curation, Investigation, Methodology, Software, Writing - original draft, review & editing.

Yuxing Tang:
Conceptualization, Formal analysis, Investigation, Methodology, Software, Visualization, Writing - original draft, review & editing.

Xiaohui Lin:
Investigation, Data Curation, Validation, Writing - review.

Rushan Ouyang:
Investigation, Data Curation, Validation, Writing - review.

Mingxiang Wu:
Investigation, Data Curation, Validation, Writing - review.

Mei Han:
Conceptualization, Investigation, Resources, Writing - review, Supervision.

Jing Xiao:

Investigation, Resources, Writing - review, Supervision.

Lingyun Huang:

Resources, Writing - review.

Shibin Wu:

Methodology, Writing - review.

Jie Ma:

Conceptualization, Investigation, Data Curation, Resources, Writing - review, Supervision.

Peng Chang:

Conceptualization, Investigation, Methodology, Resources, Writing - original draft, review & editing, Supervision.