UNIVERSITY OF TURKU

Studies on deep learning approach in breast lesions detection and cancer diagnosis in mammograms

Smart Systems

Master's Degree Programme in Information and Communication Technology

Department of Computing, Faculty of Technology

Master of Science in Technology thesis

Author:

Huairui Zhao

Supervisors:

Asst. Prof. Antti Airola

Prof. Yi Guo

June 2022

Master of Science in Technology Thesis
Department of Computing, Faculty of Technology
University of Turku

Subject: Smart Systems
Programme: Master's Degree Programme in Information and Communication Technology
Author: Huairui Zhao
Title: Studies on deep learning approach in breast lesions detection and cancer diagnosis in mammograms
Supervisors: Asst. Prof. Antti Airola, Prof. Yi Guo
Number of pages: 55 pages, 0 appendix pages
Date: June 2022

Breast cancer accounts for the largest proportion of newly diagnosed cancers in women recently. Early diagnosis of breast cancer can improve treatment outcomes and reduce mortality. Mammography is convenient and reliable, which is the most commonly used method for breast cancer screening. However, manual examinations are limited by the cost and experience of radiologists, which introduce a high false positive rate and false examination. Therefore, a high-performance computer-aided diagnosis (CAD) system is significant for lesions detection and cancer diagnosis. Traditional CADs for cancer diagnosis require a large number of features selected manually and remain a high false positive rate. The methods based on deep learning can automatically extract image features through the network, but their performance is limited by the problems of multicenter data biases, the complexity of lesion features, and the high cost of annotations. Therefore, it is necessary to propose a CAD system to improve the ability of lesion detection and cancer diagnosis, which is optimized for the above problems.

This thesis aims to utilize deep learning methods to improve the CADs' performance and effectiveness of lesion detection and cancer diagnosis. Starting from the detection of multi-type lesions using deep learning methods based on full consideration of characteristics of mammography, this thesis explores the detection method of microcalcification based on multiscale feature fusion and the detection method of mass based on multi-view enhancing. Then, a classification method based on multi-instance learning is developed, which integrates the detection results from the above methods, to realize the precise lesions detection and cancer diagnosis in mammography.

For the detection of microcalcification, a microcalcification detection network named MCDNet is proposed to overcome the problems of multicenter data biases, the low resolution of network inputs, and scale differences between microcalcifications. In MCDNet, Adaptive Image Adjustment mitigates the impact of multicenter biases and maximizes the input effective pixels. Then, the proposed pyramid network with shortcut connections ensures that the feature maps for detection contain more precise localization and classification information about multiscale objects. In the structure, trainable Weighted Feature Fusion is proposed to improve the detection performance of both scale objects by learning the contribution of feature maps in different stages. The experiments show that MCDNet outperforms other methods on robustness and precision. In case the average number of false positives per image is 1, the recall rates of benign and malignant microcalcification are 96.8% and 98.9%, respectively. MCDNet can effectively help radiologists detect microcalcifications in clinical applications.

For the detection of breast masses, a weakly supervised multi-view enhancing mass detection network named MVMDNet is proposed to solve the lack of lesion-level labels. MVMDNet can be trained on the image-level labeled dataset and extract the extra localization information by exploring the geometric relation between multi-view mammograms. In Multi-view Enhancing, Spatial Correlation Attention is proposed to extract correspondent location information between different views while Sigmoid Weighted Fusion module fuse diagnostic and auxiliary features to improve the precision of localization. CAM-based Detection module is proposed to provide detections for mass through the classification labels. The results of experiments on both in-house dataset and public dataset, 0.92@0.52 and 0.96@0.77 (recall rate@average number of false positive per image), demonstrate

MVMDNet achieves state-of-art performances among weakly supervised methods and has robust generalization ability to alleviate the multicenter biases.

In the study of cancer diagnosis, a breast cancer classification network named CancerDNet based on Multi-instance Learning is proposed. CancerDNet successfully solves the problem that the features of lesions are complex in whole image classification utilizing the lesion detection results from the previous chapters. Whole Case Bag Learning is proposed to combined the features extracted from four-view, which works like a radiologist to realize the classification of each case. Low-capacity Instance Learning and High-capacity Instance Learning successfully integrate the detections of multi-type lesions into the CancerDNet, so that the model can fully consider lesions with complex features in the classification task. CancerDNet achieves the AUC of 0.907 and AUC of 0.925 on the in-house and the public datasets, respectively, which is better than current methods. The results show that CancerDNet achieves a high-performance cancer diagnosis.

In the works of the above three parts, this thesis fully considers the characteristics of mammograms and proposes methods based on deep learning for lesions detection and cancer diagnosis. The results of experiments on in-house and public datasets show that the methods proposed in this thesis achieve the state-of-the-art in the microcalcifications detection, masses detection, and the case-level classification of cancer and have a strong ability of multicenter generalization. The results also prove that the methods proposed in this thesis can effectively assist radiologists in making the diagnosis while saving labor costs.

Key words: detection in Mammography, deep learning, weakly supervised learning, feature fusion, multi-view information enhancing.

Table of contents

1 Chapter 1: Introduction

1.1 Background

Breast cancer accounts for around 31% of newly diagnosed cancers in women worldwide and is the second leading cause of cancer-related deaths among women globally[1]. In China, breast cancer is also the most common cancer diagnosed in women and there will be 429,105 new cases in 2022 by estimation[2]. Fortunately, early detection of breast cancer has been proven to enhance treatment results and reduce mortality[3-4]. The 5-year survival rate of patients who have early treatments is significantly higher than that of patients with advanced disease, which can reach more than 90%[5].

Breast tissue biopsy is a clinical method for diagnosing breast cancer, and its results are regarded as the ground truth for cancer diagnosis. However, the biopsy is invasive and not suitable for screening. Medical Imaging, which can present the suspicious lesions associated with cancer, such as microcalcification and mass, is an ideal method for breast cancer screening because of its non-invasiveness, independence, and high diagnostic performance. Therefore, imaging is a popular modality presently used for breast cancer screening[6]. Imaging generally includes ultrasonography, breast magnetic resonance imaging (MRI), and Mammography[7], and can present the suspicious lesion.

Ultrasonography: The use of ultrasound imaging to diagnose breast cancer is an important method[8]. However, most microcalcifications are not detectable by breast ultrasonography[9].

Breast MRI: It is not commonly used for routine screening because of its high cost. When mammography and ultrasonography fail to diagnose suspicious nipple discharge, MRI may be employed[10].

Mammography: The mammography can better display microcalcifications with high imaging resolution while it is inexpensive and easy to be implanted. Therefore, it is recommended as the primary method for screening early breast cancer in current guidelines[11].

Currently, mammographic examinations are mainly performed by radiologists by reading mammograms. However, the different experiences of radiologists, the decreased attention due to the prolonged reading, and the lesion features that are difficult to detect in dense breast tissue result in errors in diagnosis outcomes. In these false examinations, benign lesions can

be misinterpreted as cancer (false positive), while malignancies may be ignored (false negative). As a result, radiologists fail to detect about 15% of breast cancers, which more likely appears in cases of women with dense breasts[12]. One of the solutions is blinded double reading, which improves the cancer detection rate. However, double reading is not a cost-effective choice, according to multiple studies[11-14].

Traditional computer-aided diagnosis (CAD) systems can be regarded as an alternate framework that functions as a second reader to improve interpretation performance. According to research in the early years, using traditional CAD systems increase the efficiency in the detection of cancer by improving performance in detecting microcalcification and mass in mammograms[15-16]. Instead of pre-taught algorithms used by traditional CADs, the new generation of CADs systems based on deep learning can integrate cancer features that reflect underlying pathophysiology and are difficult to be found by the naked eye in images. More research on CAD based on deep learning has emerged in recent years and achieved great performance[17]. Thus, a high-performance deep learning approach for breast lesions detection and cancer diagnosis in mammography is significant for the reduction of radiologists' workload and improvement of diagnostic accuracy.

1.2 Features of Mammography

Mammography uses a low-dose X-ray to make projections of the organizational structure of the breast in different views as shown in Figure 1.1. Generally, there are two projection views in standard mammography: the craniocaudal (CC) view is a projection that displays the medial and external lateral portions of the breast, while the mediolateral oblique (MLO) view is the most important projection taken from the oblique view. Because the different lesions or tissues appear in different grayscales after processing the projection signal in mammography, the lesions or cancer can be detected in mammograms. The two most significant signs of malignancy in mammograms are microcalcifications and masse.

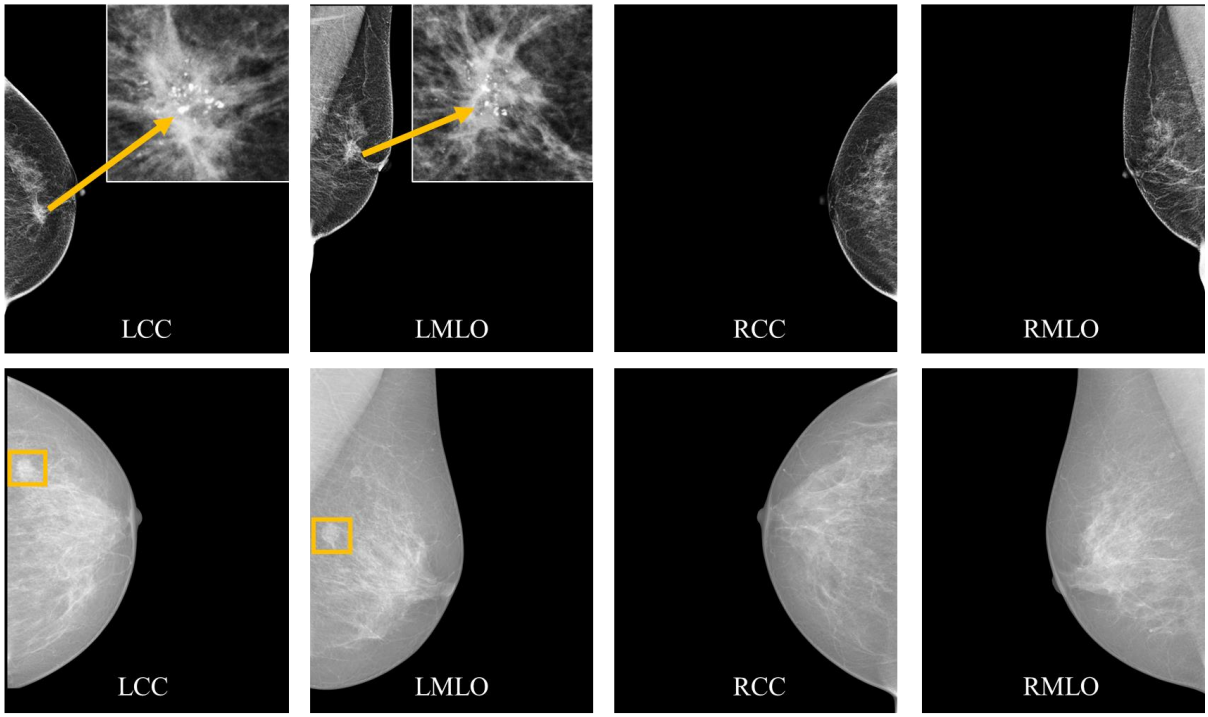

Figure 1.1 Demonstration of mammograms and breast lesions. The first and second rows show two cases with microcalcification and mass, respectively.

When utilizing the deep learning approach for lesion detection or cancer diagnosis in mammography, it is important to fully consider the characteristics of mammograms and lesions, and develop appropriate methods to assist radiologists to diagnose. The main characteristics and related issues of mammograms and lesions are as follows:

Microcalcification: Breast microcalcifications are specks of calcium scattered in the breast and generally appear as white high-density tissue in mammograms. Different types of microcalcification (benign or malignant) need to be distinguished by distribution and morphology[18]. More than 30% of breast cancer are associated with microcalcifications according to the study, and this proportion will increase to more than 70% when only considering cancer in situ[19]. However, microcalcifications generally have poor contrast with surrounding breast tissue, especially in young women[20]. Benign microcalcifications are extremely tiny while malignant microcalcifications are scattered in tissue. It is not very easy to find and distinguish these microcalcifications with the naked eye, which brings difficulties to the timely detection and diagnosis in clinical.

Mass: Breast mass is a lump in the breast. In mammograms, radiologists diagnose whether the mass is benign or malignant based on the shape, margin, and density of the mass[18]. It is difficult to find the border between the masses and normal tissue. because breast masses have similar and ambiguous characteristics with the normal tissue and could even be occluded by

normal tissue in some cases. Due to the above reasons, mass detection in mammograms is harder than microcalcification detection, which is verified by the studies[21].

High Resolution: Mammograms have low noise and high-resolution, close to 4000*3000 with 12 million pixels, compared to ultrasound images which contain speckle noise that destroys image quality[22]. However, high-resolution images with small objects, such as some microcalcifications that are only about 50 pixels, are the challenging task for deep learning-based image detection and classification. One reason is that the computing resources and the inference times of CAD greatly increase as the size of the image increases, which depends on the architecture of the deep learning model. The other reason is the detection performance of small-size objects is far from satisfactory[23].

Projection View: The principle of mammography leads to some situations that lesions, especially breast masses, are occluded by other breast tissue. In another aspect, multi-projection views provide compensation information for lesions.

Multicenter Biases: Different clinical settings (imaging protocol, scanner, etc.) of mammography provide multicenter biases in mammograms. The most significant bias among multiple centers is the changes in intensity values, which leads to the different contrast between the fibroglandular tissues and the adipose region of the breast in different mammograms. According to the study, more than 94% of published studies in AI CAD of medical images did not include multicenter validation, which is the reason that these studies are difficult to deploy in changing clinical applications[24].

1.3 Related Works

It is necessary to fully consider the above characteristics of mammograms and lesions, and find the targeted solutions when studying on deep learning approach for lesion detection and cancer diagnosis. The cancer-related lesions in mammograms are complex as shown, which are multiple classes objectively. Thus, it is hard to directly perform binary classification on such images in deep learning. The idea of this thesis for the above problems is to separate the direct classification into two steps: lesions detection and cancer classification based on the detection. The related works about each part are as follows.

1.3.1 Microcalcification Detection

The traditional CAD of microcalcification follows similar image processing steps: images augmentation, extraction of regions of interest (ROI), features calculation, and classification. To identify microcalcification, the early method clustered the area, based on the feature of appearances, such as shape and gray, by setting a gray threshold[25]. However, this method faced a high false positive rate because the gray threshold for segmentation is hard to be selected and features used for classification are simple. With the development of machine learning, researchers use different classifiers, such as decision trees, support vector machines (SVM), and artificial neural networks, to map features to higher-dimensional spaces for classification[26-27]. Among these classifiers, SVM is the most widely used[28]. Phadke et al. used SVM to classify manually ROI of abnormal tissues by fusion features consisting of local features and global features[29]. However, traditional image processing ROI selection is not robust and manual features design should select the best features among many features by performing lots of experiments.

In recent years, deep learning has been widely used in medical image processing[30]. Compared with traditional methods, deep learning uses the convolutional neural network (CNN) and its variants to automatically learn and extract the most relevant features. Zhang et al. proposed a cascaded network that included a false positive reduction network for microcalcifications[31]. Xi et al. calculated class activation mappings (CAMs) to create a heatmap, which plays a role in localizing microcalcifications[32]. Both two works regard detection as a process involving multiple stages, which are difficult to optimize for complex tasks. Ayelet et al. first presented an integrated detection network based on Faster R-CNN that combines positioning and classification, and a cascade classifier is added to the end to further reduce the false positive rate of the network[33]. Cao et al. adopted a segmentation network to detect individual microcalcifications. A 512*512 sliding window was used to perform detection on the whole mammogram, but clustered microcalcifications exceeding the size of the chosen slice were not included in the training stage[34]. In the above works, various features of multiscale objects limit the performance of their methods.

1.3.2 Mass Detection

Full-supervised deep learning methods are the most commonly utilized in CAD because of their high performance[35]. However, the collection of high-quality annotations that

contributes to the high performance is time-consuming. Three weakly supervised methods are commonly used to alleviate the lack of annotations, including transfer learning, multiple instance learning, and class activation mapping[35-41]. Transfer learning approaches fine-tuned the fully-supervised detection model to the dataset with weak annotation[37-38]. However, the supervised training stage still needs a large number of lesion annotations. Multiple instance learning (MIL) provides supervision for the entire bag instead of its instances. That is, instead of patch-level labels, the MIL approach for lesion detection simply requires image-level labels. Shen et al. proposed a Globally-Aware Multiple Instance Classifier with joint training of patches and whole images using image-level labels[39]. The performance is constrained by the loss design and the additional patches selection algorithms, which leads to a high false positive rate. Class Activation Mapping (CAM) techniques are innovated for the visual interpretation of CNN and applied in weakly-supervised detections in natural images[40]. Liang et al. utilized the self-training strategy in CAM for localization of breast masses[41]. However, the self-training strategy is difficult to find hard samples that occlude by healthy breast tissues which have similar grayscale to the mass.

Some fully-supervised learning works focus on how to provide more additional information utilizing multi-view images. In early approaches, they do not provide end-to-end frameworks and the fusion step is applied after first stage detection[42-43]. Two-stage design is a disadvantage because the second stage is sensitive to the detection results and hard to be optimized. Ma et al. proposed an end-to-end model including a relation fusion module, where the extra cross-view information was extracted from patches[44]. However, the proposal of patches affects its performance easily. Liu et al. utilized the graph convolution network (GCN) to extract correspondence in multi-view mammograms mapped by nodes[45]. But the region-level correspondence extracted by the GCN is not precise due to the tissue deformation in the projection process. Inspired by the above works, we apply the idea of multi-view in the weakly supervised detection.

1.3.3 Cancer Diagnosis

The cancer diagnosis of CADs in mammography aims to help radiologists to determine the lesions in mammograms as benign or malignant. The deep learning methods for cancer diagnosis generally can be divided into two categories based on whether the input is ROI of lesion or is the entire mammogram.

One is the ROI-based classification, of which inputs are ROIs of abnormalities. Arevalo et al. proposed a mass classification network based on deep learning, which utilizes the SVM to follow the network to finetune the classification while Qiu et al. used the end-to-end CNN to realize the classification of mass patch[46-47]. Sun et al. developed a semi-supervised deep learning network, of which the accuracy can be improved by using a large amount of unlabeled data[48]. Morrell et al. proposed a fully convolutional network and deformable convolutional nets to improve the performance of CAD for mammography[49]. However, the above method based on ROI focuses only on binary classification of certain lesions, which faces the problem of the acquisition of lesion patches. The performance of these methods would be limited by the accuracy of the patch selection and obtaining accurate ROI patches manually is time-consuming. Thus, we consider taking another approach: the classification based on the whole image, of which inputs are whole images. In this type of approach, the binary classification indicates that the images contain or do not contain lesions associated with cancer.

A binary classification that is directly performed on the whole image is easy to cause insufficient consideration of various lesions associated with cancer in mammograms. Thus, most works try to add lesion information in whole image classification. Carneiro et al. integrated six CNN models for multi-view mammograms with and without lesion masks to enhance the performance of the whole image classification for one side breast[50]. Dhungel et al proposed a multi-view deep residual neural network named mResNet for the classification of mammograms, which also concatenates the six images features in one fully connected layer[51]. Zhao et al. proposed a bi-lateral attention module and bi-projection attention module, which are more refined in designing the modules of multi-view relationships[52].

To optimize the feature extraction of whole image classification in the network, Shen et al. utilized the ROI patch pretraining the classification network while Shanms et al. simultaneously trains a GAN network that is trained on lesion patches and share a feature extraction backbone with the classification network[38,54]. Hu et al proposed a weighted MIL network with multiple gamma-corrected images input[55]. The above works focus only on the direct classification of one side breast or one image instead of the whole case. McKinney et al proposed an AI system to realize the whole case classification and make a comparison with radiologists. This AI system consists of three deep learning networks for lesion level, breast level, and case level classification respectively, which diagnose according

to the average score of three networks[56]. However, this scoring mechanism is not optimal for model integration.

According to the above overview of works related to lesion detection and cancer diagnosis, some main challenges are still faced in current methods: Diversity of features of multiscale microcalcification in microcalcification detection; Lack of localization information in mass weakly supervised detection; Complexity of lesion features in whole image classification. Our methods are introduced next, which are proposed aiming to address these challenges.

1.4 Our Work and Contributions

In this thesis, we successfully develop a computer-aided diagnosis that can efficiently use existing clinical annotations to achieve the detection of multi-type lesions and help radiologists make the diagnosis in clinical.

Figure 1.2 is the outline of the thesis. Firstly, stating from lesions detection, we propose a network for detection of microcalcification in mammography based on multiscale features fusion. Secondly, considering compensation information from multiple views, we propose a weakly supervised multi-view enhancing network for mass detection in mammography. Finally, utilizing the works of lesions detection motioned above, we propose a breast cancer classification network based on multi-instance learning, which achieves high-performance cancer diagnosis.

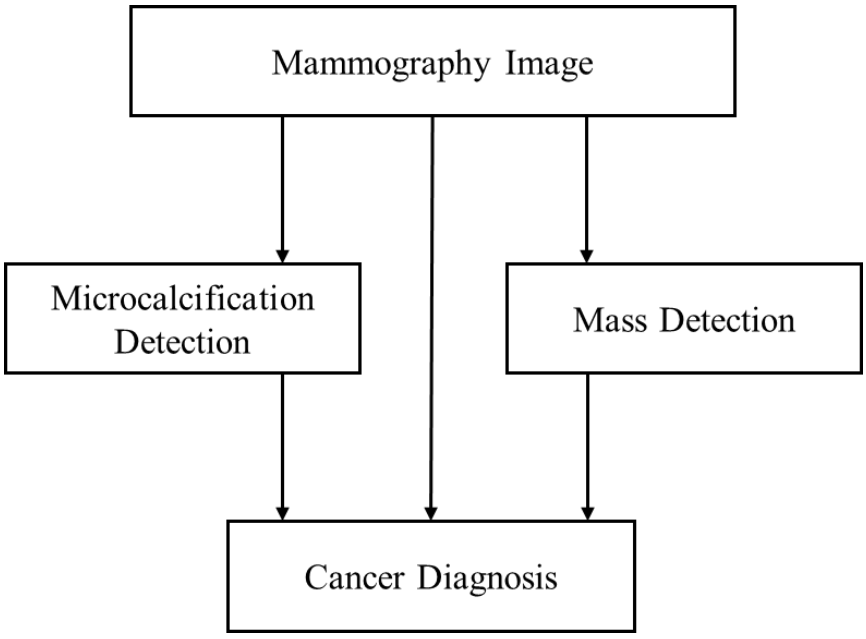

Figure 1.2 Outline of the thesis work.

In the above outline of thesis work, the main contributions of this thesis can be summarized as follows:

Aiming to overcome the problems of multicenter data bias, the high-resolution of mammograms, and scale differences between microcalcifications, a pyramid feature network for microcalcification detection (MCDNet) is proposed. In MCDNet, the Adaptive Image Adjustment mitigates the impact of multicenter biases and maximizes the input effective pixels. Then, the proposed shortcut connection pyramid network makes sure the detection ability of the network for multiscale objects. Based on the structure, Weighted Feature Fusion is proposed to improve the detection performance of both scale objects by learning the contribution of feature maps in different stages. The experiments show that MCDNet is more precise and robust than other methods in microcalcification detection, which can effectively help radiologists identify two types of microcalcifications in clinical applications.

Aiming to solve the problem of the lack of lesion-level labels, a weakly supervised multi-view enhancing mass detection network (MVMDNet) is proposed, which can be trained on the image-level labeled dataset to detect breast masses. The image preprocessing including Adaptive Image Adjustment makes sure the alignment of multi-view and reduces the biases. Multi-view Enhancing module is proposed to extract correspondent localization information between different views, and enhance these features by fusing diagnostic and auxiliary feature maps, which solve the lack of localization information in weakly supervised training with only image-level labels. CAM-based detection module is proposed to provide detections for mass through the enhanced localization information from Multi-view Enhancing modules. The experiments conducted on both in-house and public datasets demonstrate that MVMDNet achieves state-of-art performances and a robust generalization among weakly supervised methods.

Aiming to face the complex features of multiple lesions in mammograms, a breast cancer diagnosis network based on Multi-instance Learning named CancerDNet is proposed to realize the full images classification for the whole case, which utilizes the lesion detection results from the previous chapters. The Whole Case Bag Learning is proposed to combined the features extracted from four-view mammograms and works as a radiologist to realize the classification of each case. Low-capacity Instance Learning and High-capacity Instance Learning successfully integrate the detections of multiple types of lesions, which are the works of chapters 2 and 3, into the CancerDNet, so that the model can fully consider lesions

with complex features in the classification task. The experiments conducted on both in-house and public datasets show that CancerDNet achieves a high-performance cancer diagnosis.

1.5 Datasets and Implementation

The datasets we used in this thesis include one in-house dataset and one public dataset. The in-house dataset is collected from Renji Hospital, Shanghai Jiao Tong University School of Medicine, which contains a total of 1491 cases. Informed consent was obtained from all individual participants included in the study. This study has been approved by the ethics committee of the Renji Hospital. The public dataset is INbreast, which contains a total of 115 cases, and its usage has been authorized by the publisher[57]. Both these two datasets are full-field digital mammography. However, the different clinical settings of these two datasets provide multicenter biases in mammograms. The most significant bias is the changes in intensity values, which leads to the different contrast between the normal and the abnormal tissues of the breast. As shown in Figure 1.1, the mammogram in the second row is from INbreast, where the contrast between mass and normal region is lower than that in the mammogram of the in-house dataset which is demonstrated in the first row. The details of the use of the datasets are described in each chapter.

All experiments of this thesis are implemented on the deep learning framework, TensorFlow (https://www.tensorflow.org/), PyTorch (https://pytorch.org/), etc.

1.6 Thesis Structure

Chapter 1 is the introduction that introduces the relevant background of breast cancer diagnosis and the significance of deep learning-based computer-aided diagnosis in mammography screening. It also describes the related work and contributions of this thesis. Finally, the datasets and implementation platform are briefly introduced.

Chapter 2 focuses on the detection of microcalcification, which is one of the important indicators of breast cancer. First, the development background of MCDNet is introduced, and then each contribution is explained in detail. The experiments that validate the performance of each contribution and whole network are shown with a discussion of the results.

Chapter 3 introduces the detection of breast mass in mammography. Similarly, the background of how to solve the problem of the leak of labels is stated. Then, the innovation details of MVMDNet are explained with a detailed network structure. Finally, ablation studies

and comparison experiments are presented to demonstrate the performance of MVMDNet, and interpretation of the results is discussed.

Chapter 4 introduces the CancerDNet, which is a cancer classification network based on the works of the previous two chapters. Firstly, the idea that how to realize the whole image classification is explained. Secondly, the methods of CancerDNet are illustrated with diagrams. Finally, comparative experiments are implemented, and a discussion of the above comparison results is presented.

Chapter 5 is the conclusion and perspective. It summarizes the whole works of the thesis and proposes targeted improvement directions for some optimizable parts.

2 Chapter 2: Microcalcification Detection in Mammography Based on Multiscale Features Fusion

2.1 Introduction

Lesions detection can help the cancer diagnosis network focus on the regions or features, which is needed to consider when classifying mammograms. This chapter focuses on the detection of microcalcifications in mammograms. Mammograms generally display two types of microcalcifications. One appears as an amorphous clustered group and is strongly related to cancer. The other looks like tiny and individual points and is generally found in benign tumors[18]. High-precision detection for individual and clustered microcalcifications in mammography is important for the early diagnosis of breast cancer. However, individual microcalcifications are observed on a much smaller scale than clustered microcalcifications and the contrast between the microcalcification and the background tissue is generally low[20]. Therefore, it is time-consuming for radiologists to identify these microcalcifications, which introduces some error diagnoses.

Some examples of microcalcifications are shown in Figure 2.1, images (a)-(c) are collected from the in-house dataset with the lesions labeled by radiologists using green bounding boxes. Image (d) is collected is from the INbreast dataset. Image (a) contains a large region of clustered microcalcifications while (b) contains a small region of clustered microcalcifications. Image (c) and (d) contain extremely tiny individual microcalcifications.

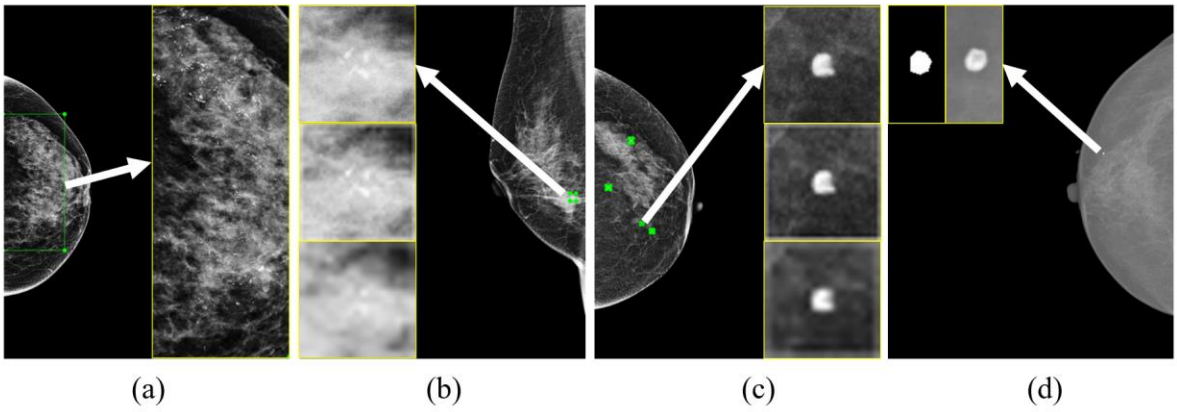

Figure 2.1 Illustrations of microcalcifications. (a) and (b) demonstrate clustered microcalcifications. (c) and (d) demonstrate individual microcalcifications.

A computer-aided detection (CAD) system can help radiologists improve efficiency and accuracy, which could detect lesions automatically. Widely used deep learning approaches

and their variants can automatically learn and extract the most relevant features[30]. However, the performances of related methods[31,34] are limited by several challenges. First, the performance of deep learning models is generally sensitive to the dataset and is limited by multicenter biases. As shown in Figure 2.1 (c) and (d), the contrast between microcalcifications and normal tissues are different, which results in poor robustness when conducting the model on other datasets. Second, the input images are generally resized to a low resolution due to the constrain of computational resources. Figure 2.1 (b) and (c) demonstrate the appearances of lesions presented in different resolutions, where the resolutions of patches in yellow boxes from top to bottom are original, 1/4, and 1/8, respectively. The edge features of small objects and the internal features of clustered microcalcifications are destroyed by the down resize operation and further lost during the convolutions of downsampling. A commonly used method is to crop the image into patches, which fail to detect the large-region clustered microcalcifications, such as image (a) in Figure 2.1[33-34]. Third, the traditional CNN is weak for the detection of multiscale objects, such as the microcalcifications in this work. After feature propagation in CNN, the localization information and internal features are not precise. As a result, small objects cannot be detected while large objects are scattered.

We propose a microcalcification detection network MCDNet to solve the above challenges, which realizes the precision detection of two types of microcalcification. Firstly, a built-in image preprocessing named Adaptive Image Adjustment (AIA) is proposed to mitigate the multicenter biases. Second, a feature extraction and fusion structure, Shortcut Connection Pyramid Network (SCPN), is proposed to improve the propagation of the multiscale critical features more efficiently compared with the traditional feature pyramid network (FPN)[58]. Finally, a Weighted Feature Fusion (WFF) strategy is designed in SCPN to learn the contributions of feature maps that contain different scale information.

2.2 Detection Based on Multiscale Features Fusion

2.2.1 Overview

MCDNet consists of three parts. Firstly, AIA adjusts the input mammograms to leverage the impact of multicenter biases and maximize the effective pixels of the input. Secondly, the outputs of AIA are fed into SCPN, where different scale feature maps are fused by trainable weights of WFF to improve the ability of multiscale feature extraction. Next, two branches of

CNN predict the classes and generate precise boxes on the proposed region by the region proposal network (RPN). Finally, the detection results are presented on the origin mammograms. Figure 2.2 demonstrates the framework of MCDNet.

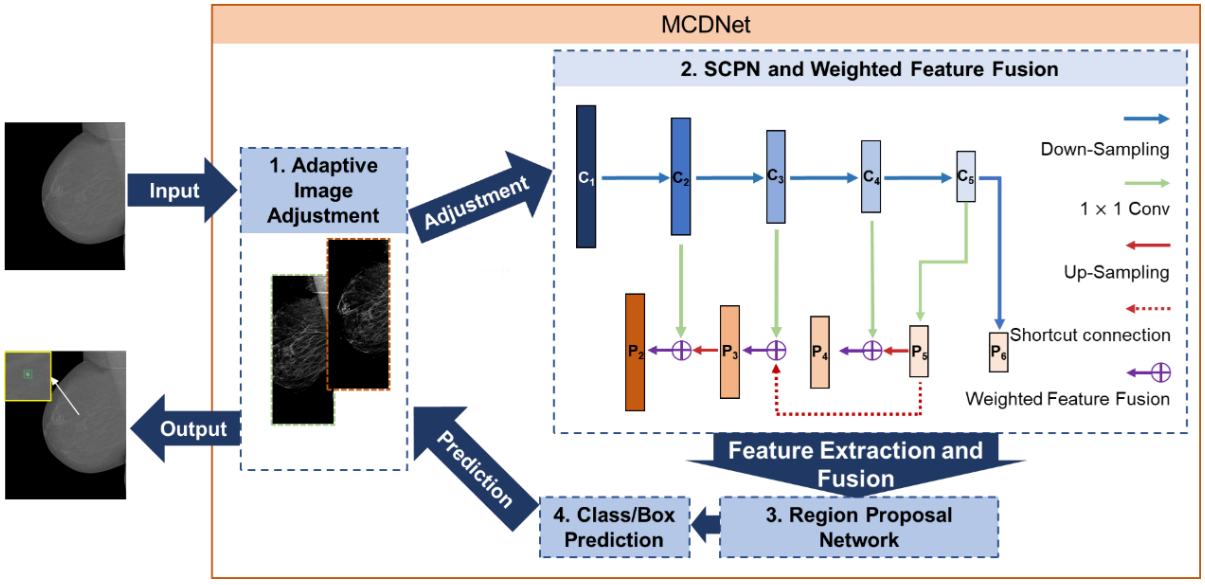

Figure 2.2 The framework of MCDNet.

2.2.2 Adaptive Image Adjustment

As shown in Figure 2.3, AIA consists of standard histogram specification and adaptive image cropping.

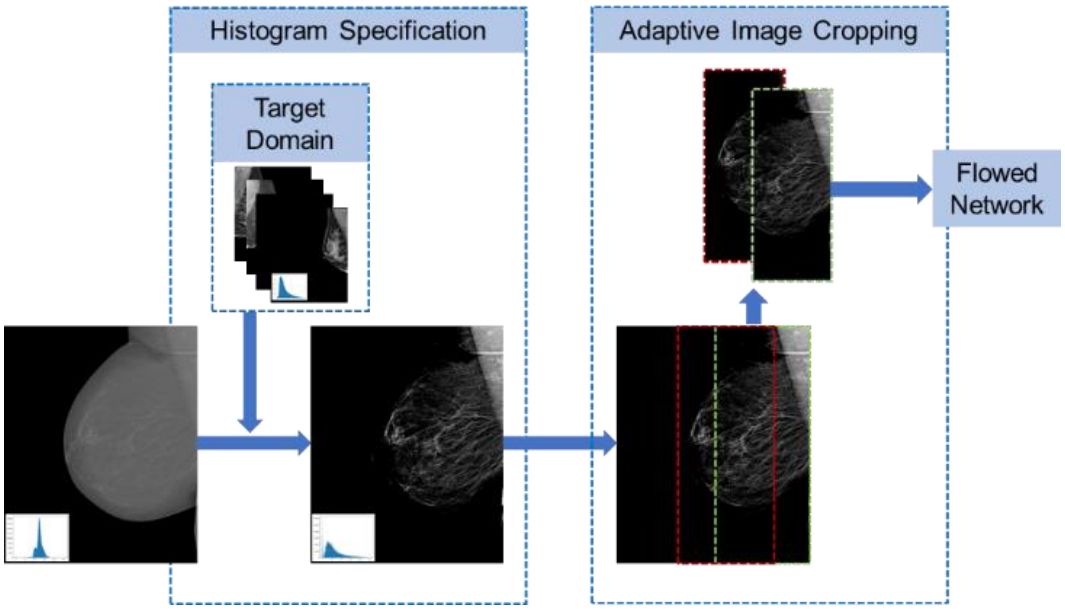

Figure 2.3 A demonstration of Adaptive Image Adjustment.

To ensure that the training and testing sets share a similar intensity distribution, a histogram specification is applied to match the histogram of different datasets. In the study, the public dataset is regarded as source mammograms while our in-house dataset is regarded as target mammograms. The cumulative distribution function (CDF) of the reference image $F_S(\cdot)$ and target image $F_T(\cdot)$ is calculated at first. Then, the gray level G_T for which $F_s(G_s) = F_T(G_T)$ is found, so that the matching function is $S(G_s) = G_T$. If the values of the CDF are discrete, the nearest $F_s(G_s)$ is to be found. Finally, $S(\cdot)$ is applied to each pixel of the reference mammogram.

Many ineffective pixels of full mammograms are useless for the detection of the object and waste training resources so that the non-breast region should be removed. Previous works which crop mammograms into several patches are proved to negatively impact the detection. Cropping according to a minimal rectangle is a simple and frequent strategy in other studies. However, it results in varied sizes, which causes image distortion when the fixed input size is used by the deep learning network.

To mitigate this effect, we explore novel adaptive cropping rules implemented in AIA on the network's input and output ends. Because most breasts exist on the one side of mammograms, the picture is cropped from the middle in AIA. The cropping window automatically shifts if the breast region is more than half of the image. If the breast extends beyond the midline, as illustrated in Figure 2.3, the cropping window marked by green automatically shifts to the left to guarantee that the left edge of the cropping window marked by red is aligned with the breast's edge, and the size remains intact. Both cropped mammograms are sent into the training or detection queue at the same size.

2.2.3 Shortcut Connection Pyramid Network

Clustered microcalcifications in mammograms are typically one hundred times bigger than individual microcalcifications. Because clustered microcalcifications generally are dispersed throughout the breast, extra semantic information is required to retain the object's integrity during the detection. On the contrary, the localization of individual microcalcification necessitates additional edge features and shallow information from high-resolution feature maps.

It's challenging to preserve valuable information on the same feature map layer level for both types of microcalcifications with distinct sizes at the same time for a generic CNN detection

network. Low-level features have more edges and accurate localization information during downsampling while high-level features have more spatial semantic information. We propose a multiscale feature pyramid architecture with a shortcut connection named SCPN inspired by the FPN to solve the multiscale object detection challenge. Feature maps are sampled and fused following the paths in SCPN to generate multiscale feature maps for prediction. Especially along with the shortcut connection that crosses over one layer, the required information is effectively propagated.

This traditional pyramid structure is expressed as $P_i = Conv(C_i + Resize(P_{i+1}))$. In the formula, P_i is the feature map in the upsampling and fusion process on which the RPN runs while C_i is the feature map in the downsampling process as shown in Figure 2.2. Starting from C_4, the typical pyramid network prioritizes semantic and spatial features above edge features, as seen in Figure 2.4. Because of the addition of high-level feature maps C_i, we think that the high-level prediction feature maps P_i include more worthless information for individual microcalcifications, which has detrimental impacts on localization.

It is difficult to prediction feature maps that aliases numerous high-level feature maps, such as P_3, to get correct localization information in the traditional pyramid structure. As the comparison between Figure 2.4 (a) and (g), the localization of the individual microcalcification is not precise.

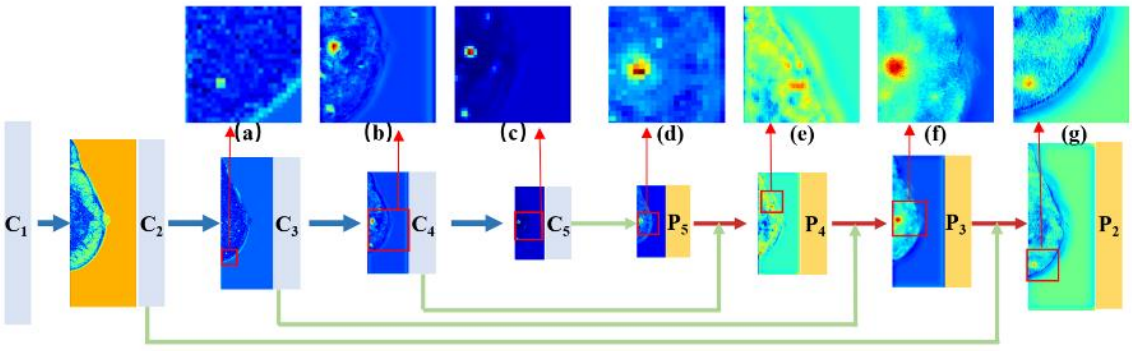

Figure 2.4 The heatmaps in the traditional pyramid structure. (a)-(g) are feature maps.

Figure 2.4 (f) shows the imprecise localization of individual microcalcifications in the P_3 layer. Our intuition is that minimizing the amount of confusing and useless information stacked in the P_3 could enhance the localization precision of individual microcalcification. Therefore, we design a cascaded information propagation shortcut connection structure.

$$P3 = Conv(C_3 + Resize(P_5)) \quad (2.1)$$

Formula 2.1 shows that the quantity of aliased high-level information gained through the proposed shortcut connection over P_4 is minimized. That is, the SCPN can minimize the high-level feature maps contributions when the information flow is propagated in the network, such as C_4, P_5, C_5, P_6, and enhance the localization precision. Sematic and spatial information from high-level feature maps, on the other hand, is critical for ensuring the entire recognition of whole lesions for clustered microcalcifications. The key contributing feature map layers P_6 to P_4 preserve the high-level feature maps containing semantic and spatial information according to the calculation of downsampling ratio.

2.2.4 Trainable Weighted Feature Fusion

The commonly used feature layer fusion approach resizes one layer of a feature map by upsampling it to the size of another layer and summits the two feature maps directly. $P_i = C_i + Resize(P_{i+1})$ is the formula describing the fusion process, where the contributions of the two input layers C_i and P_{i+1} are equivalent. However, as previously indicated, feature maps of different resolutions contain different information. As a result, these feature maps' contributions are unbalanced. The localization and edges information contained in the low-level feature maps is important to the individual microcalcification, whereas the spatial semantic information in the high-level features serves primarily as a supplement. On the contrary, semantic and spatial information are critical for clustered microcalcifications. Thus, the Weighted Feature Fusion approach is employed to solve this problem. Trainable weights are used to fuse two feature maps based on the parameter w_i, rather than explicitly adding two feature maps. This technique is carried out using Formula 2.2 during the training stage:

$$Out = \sum_i \frac{max(0,w_i)}{\epsilon+\sum_j max(0,w_j)} i \tag{2.2}$$

in which $max(0, w_i)/\left(\epsilon + \sum_j max(0, w_j)\right)$ is the weight assigned to each feature map, i, in the fusion stage. The weights are quickly normalized to avoid potential training instability and their sum equals 1, which is the same as the outcome of a simple addition operation. A rectified linear unit (ReLU) is applied on w to ensure that $w > 0$ in the training stage. $\epsilon = 0.001$ improves the stability in the training stage. Formula 2.3 is used to explain the fusion process of P_3 :

$$P_3 = Conv\left(\frac{w_1 \cdot C_3 + w_2 \cdot Resize(P_5)}{w_1 + w_2 + \epsilon}\right) \tag{2.3}$$

2.3 Experiments

2.3.1 Dataset and Implementation

The datasets we used in the experiments include a part of the in-house dataset and INbreast dataset which contains microcalcification detection labels, as shown in Table 2.1. The in-house dataset contains two sub-datasets, where the labels include the bounding boxes of objects and the classification of objects, which are labeled by radiologists. The first sub-dataset comprises 125 mammograms from 65 cases, with 91 clustered microcalcification labels that indicate the malignant microcalcifications, and 226 individual microcalcification labels that indicate the benign microcalcifications. The second sub-dataset includes 598 mammograms from 152 cases, as well as 1672 benign microcalcification labels.

100 mammograms are collocated from INbreast, which contains 92 individual microcalcifications and 18 clustered microcalcification labels. All of the labels are derived from masks in INbreast dataset files.

Table 2.1 Overview of the datasets.

Types of microcalcification	INbreast	In-house Dataset	
		First Sub-dataset	Second Sub-dataset
Benign	92	226	1672
Malignant	18	91	-

We merged the first sub-dataset with the public dataset to create a new mixed dataset. In each time of a 5-fold cross-validation strategy implemented in experiments, a total of 180 mammograms from the mixed dataset were utilized for training and 45 mammograms for testing. In addition, the model's performance of the individual microcalcification detection was conducted on the second sub-dataset.

MCDNet was built using the TensorFlow framework and trained on a machine with an NVIDIA Tesla V100 SXM2 graphics processing unit (GPU). The momentum gradient descent approach was utilized in the training stage with a learning rate of 0.001 and a momentum hyperparameter of 0.9. After AIA, the network's input resolution was 1200*2954, and the average GPU memory utilization was 17 GB while the batch size is 1. We used cross-entropy loss and smooth L1 loss in networks followed the RPN.

2.3.2 Compared Methods

Firstly, we utilized commonly used detection networks as baselines, including You Only Live Once (YOLO)-V3[59], Faster R-CNN, and Faster R-CNN with an FPN, with input resolutions of 608*608, 600*600, and 600*600, respectively. Due to the structural constraints of YOLO-V3, the input resolution can only be approximated. To show the benefits of high-resolution (HR) inputs for our study, we used 1664*2048 mammograms for Faster R-CNN with an FPN. To show the contributions of our approaches, including AIA, SCPN, and WFF, we conducted an ablation study. The mixed dataset and the second sub-dataset were used for this section.

Secondly, MCDNet is compared with several state-of-the-art methods. Lu et al. proposed a cascade of boosting classifiers[60]. Liu et al. integrated clustering algorithm and a weighted support SVM[61]. Zhang et al. presented a cascade model consisting of an anomaly separation network and a discriminative model[31]. Akselrod-Ballin et al. cascaded a deep neural network following Faster R-CNN[33]. DeepLIMa uses the U-Net model with group normalization[34].

2.3.3 Evaluation Metrics

The mean average precision (mAP) and maximum recall rate (recall) metrics are used to compare the common detection network baselines and conduct ablation tests. We believe the prediction is true positive if the intersection over union (IOU) between the prediction and the ground truth annotation is more than 0.5. The average precision of two types of microcalcifications is used to calculate the mAP. A high mAP implies more reliable lesion detection findings, and a high recall rate indicates more objects can be detected.

We utilize the recall rate at a particular number of false positives per image (Recall@FPI) and free-response receiver operating characteristic (FROC) curves to compare with state-of-the-art models. Recall@FPI evaluates recall ability based on an average amount of false positives, which varies depending on the clinical application. In an FROC curve, the abscissa reflects the number of false positives, while the ordinate represents the recall rate. The higher curves indicate the better performance of the model. These metrics also evaluate the ability to be applied clinically.

2.3.4 Results and Discussion

The comparison between the baselines and our model is presented first in this section. Then we compare MCDNet's detection performance to that of other models. We also examine the contributions of our methods based on the experimental results. Here, individual microcalcification is denoted by IM, while clustered microcalcification is denoted by CM in Table 2.2.

Table 2.2 Comparison of different baselines and an ablation study.

Model		Mixed Dataset			Second Subdataset
		mAP Both	Recall (%) IM	Recall (%) CM	Recall (%) IM
Baseline	YOLO-V3	0.535	65.4	66.7	57.3
	Faster RCNN	0.670	58.7	64.8	51.0
	Faster RCNN + FPN	0.780	88.9	88.9	82.7
	Faster RCNN + FPN (HR)	0.882	92.6	96.3	89.1
Ablation Study	AIA + Faster RCNN + FPN	0.888	93.5	100	92.6
	AIA + SCPN	0.944	99.7	100	97.1
	AIA + SCPN + WFF	0.963	100	100	98.4

a）Baseline and Ablation Study

As shown in Figure 2.5, clustered microcalcifications are marked by red bounding boxes while individual microcalcifications are marked by green bounding boxes except in (f).

From Table 2.2, the accuracy of YOLO-V3 and Faster RCNN is poor. Individual and clustered microcalcification recall rates rise from 58.7% and 64.8% to 88.9% and 88.9%, respectively, after stacking an FPN structure on Faster R-CNN, indicating that the FPN is an effective structure for multiscale objects detection. Furthermore, as we increase the input resolution of Faster R-CNN, the recall rates rise to 92.6% and 96.3%, respectively, while more extreme small-scale microcalcifications with low contrast are detected, as shown in Figure 2.5 (b) and (c), because the structure and pixel details of microcalcifications are preserved.

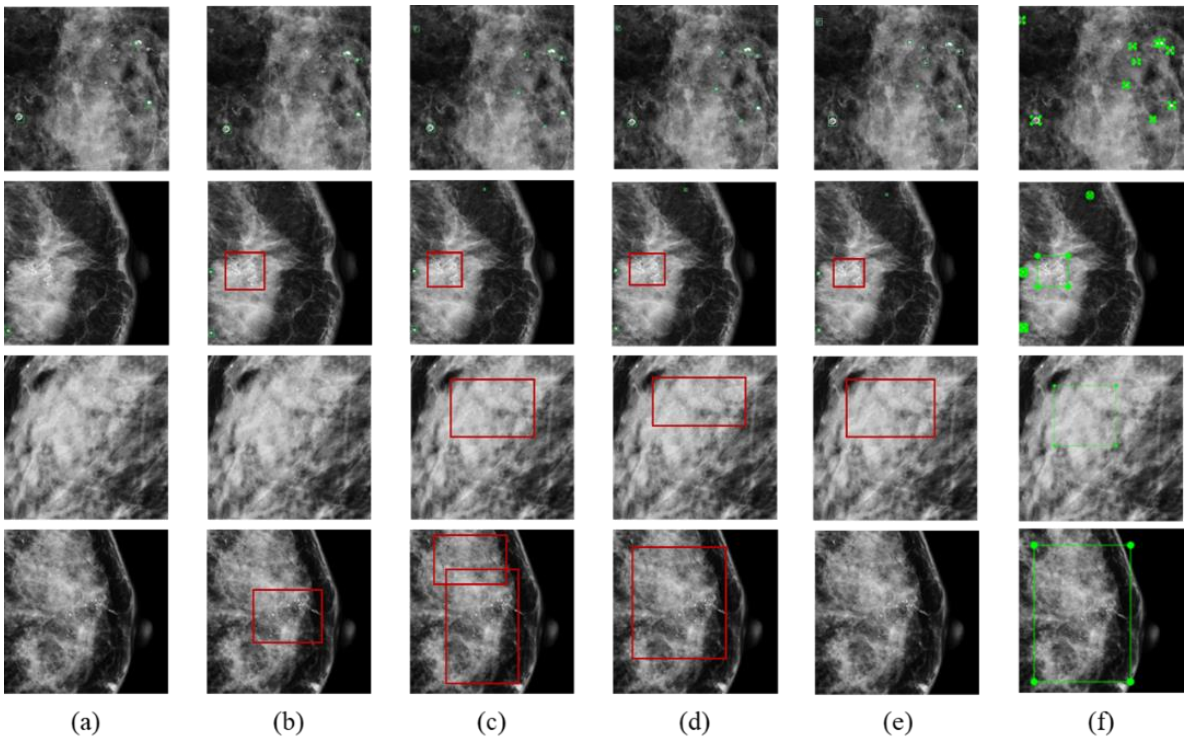

Figure 2.5 Detection results of experiments. (a)-(e) present results of the Faster R-CNN FPN, Faster R-CNN FPN with high-resolution inputs, AIA, AIA + SCPN, and AIA + SCPN + WFF, respectively. (f) is the ground truth labeled by radiologists.

Because the input resolution is increased from 1664*2048 to 1200*2954 with identical GPU memory usage, that is the number of effective pixels is doubled, the AIA approach raises the average recall rate of the two kinds from 94.5% to 96.8%. The 3.5% improvement produced by AIA in the second sub-dataset is notably substantial when compared to the 0.9% increase in the recall rate attained for individual microcalcifications in the mixed dataset. Furthermore, the model has a strong generalization ability and can mitigate the biases due to different sub-branches or multiple centers, as evidenced by the reduction in performance degradation incurred on different datasets of approximately 77%, which is a reduction of between 3.8% and 0.9%.

For clustered microcalcifications, there is a significant improvement and a 100% recall rate is attained. However, as shown in Figure 2.5 (c), some large-region clustered microcalcifications are split, and the top red bounding box is recognized as a false positive case because the IOU with the ground truth is less than 0.5. Additionally, the detection box does not correctly mark some individual microcalcifications. The mAP was lowered in these circumstances.

The mAP of the two types of microcalcifications increase from 0.888 to 0.944 when the SCPN is used, and the recall rate for individual microcalcifications increases from 93.5% to 99.7%, demonstrating that the SCPN model can solve the above problem of incorrect detection boxes. In Figure 2.5 (d), both big and tiny regional clustered microcalcifications are appropriately detected without being separated.

Another benefit of the shortcut connection structure is that it increases the model's localization ability of tiny object detection, allowing for more precise detection box regression whose IOUs are over the threshold. As shown in Figure 2.6, the heatmap generated by the SCPN is more concentrated on objects than the heatmap produced by the traditional FPN structure. In terms of propagating characteristics for localizing individual microcalcifications, we believe the SCPN is more successful than the FPN. The detection boxes are accurately regressed, allowing these objects to be properly recognized as true positives while simultaneously improving the recall rate and mAP.

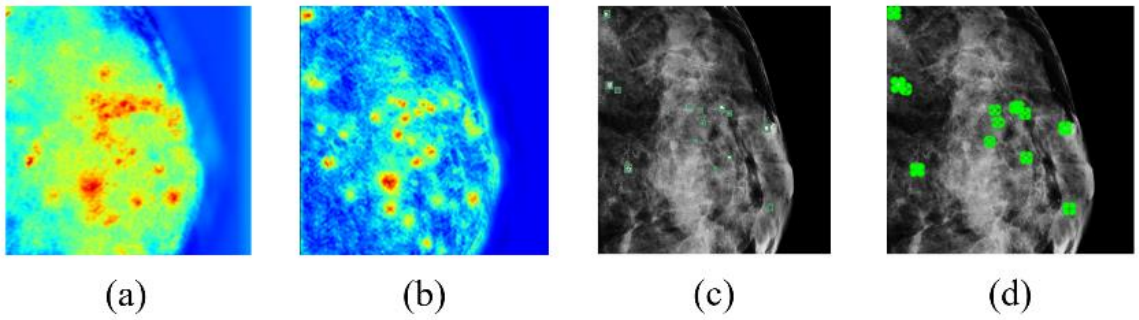

Figure 2.6 An illustration of heatmaps for different structures. (a) and (b) are heatmaps of the P_3 in the FPN and SCPN, respectively. (c) shows detection results of MCDNet. (d) is ground truth.

On our in-house dataset, weighted fusion increases the recall rate for individual microcalcifications to 100%, and the mAP increases to 0.963. The trainable weights w_i of the downsampling and upsampling feature maps produced by the trained model at various stages shows that the model tends to enhance the weights of the upsampling feature maps during the P_3 and P_4 fusion stages, and to enhance the weights of the downsampling feature maps during the P_3 fusion stage.

b) Comparison with Other Models

We compare MCDNet to other state-of-the-art models in microcalcification detection on INbreast dataset. As shown in Figure 2.7, the red, blue, and black curves are the FROC curves obtained from the detection of the individual microcalcification, the clustered

microcalcification, and both types of microcalcifications, respectively. For the two types of microcalcifications, MCDNet achieves an average of 0.84@1FPI. The FROC curves of MCDNet are higher than other models. According to these analyses, MCDNet can detect more lesions related to breast cancer than other methods under particular clinical application constraints. As shown in Table 2.3, the results of experiments conducted on INbreast also prove that the recall ability of MCDNet is better than others.

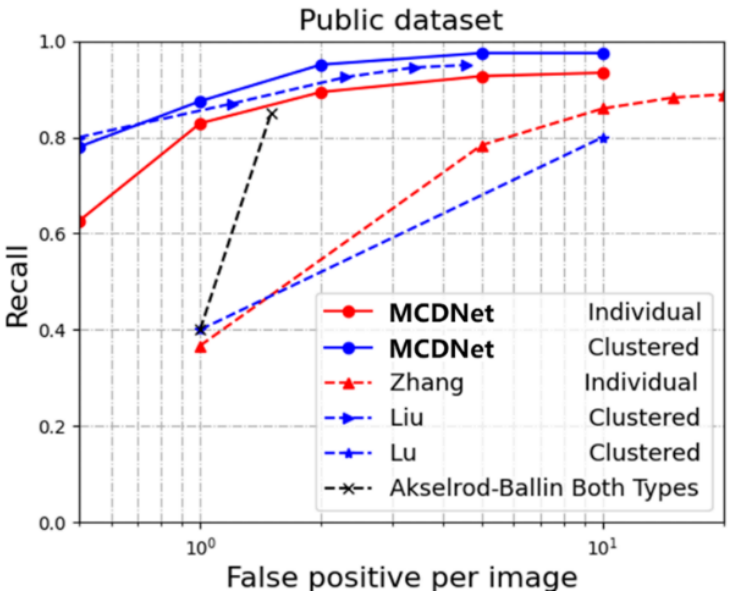

Figure 2.7 Performance on INbreast.

Table 2.3Table 2.3 Results of different methods.

Model	INbreast	In-house
Lu's	0.40@1, 0.80@10 for CM	-
Liu's	0.80@0.5, 0.95@4.6 for CM	-
Zhang's	0.89@20 for IM	0.853@1 for IM
Akselrod-Ballin's	0.85@1.5 for Both Types	0.48@1 for CM, 0.52@1 for IM
DeepIMa	-	0.912@1 for CM,0.691@1 for IM
MCDNet	0.95@2 for CM, 0.93@10 for IM	0.989@1 for CM, 0.968@1 for IM

Our model achieves the best detection performance in the in-house dataset comparison, with 0.968@1FPI for individual microcalcifications and 0.989@1FPI for clustered microcalcifications. The results in Table 2.3 show that our model outperforms the other models for both types of microcalcification. When compared to the newly proposed deep

learning model DeepLMa, our model improves microcalcification detection significantly. When the number of false positives per picture is equal to 1, our model improves a recall rate of 11.8% for both types compared to the DeepLIMa, with a considerable gain of 14.8% for individual microcalcifications.

2.4 Conclusion

In this chapter, a deep learning microcalcification detection network MCDNet is proposed to realize the detection of two types of microcalcifications. First, Adaptive Image Adjustment mitigates the impact of dataset bias and maximizes the effective pixel input. Second, Shortcut Connection Pyramid Network ensures that the feature maps retain all relevant features for multiscale objects, and improves the feature propagation efficiency. Third, the weights of each feature map in the fusion stage are optimized when training, allowing the network to learn the contributions of each feature map automatically. MCDNet achieves 96.8%@1 (individual) and 98.9%@1 (clustered) on the in-house dataset while the FROC demonstrates that MCDNet outperforms other methods in the public dataset. The results on both in-house and INbreast datasets show that MCDNet can effectively help radiologists detect and identify microcalcifications and lay a groundwork for subsequent whole-image cancer diagnosis studies.

3 Chapter 3: Weakly Supervised Mass Detection in Mammography Based on Multi-view Enhancing

3.1 Introduction

Breast mass is another suspicious lesion that can indicate breast cancer, which is tougher to be detected by CAD than microcalcification[21]. However, the high performance of the full-supervised deep learning is contributed by the high-quality annotations of lesions, whose collection is time expensive. The clinical datasets collected are generally only accompanied by diagnostic reports, which is a rough textual description of the mammograms. We face the lack of labels problem in our in-house dataset of which labels only indicate whether there is a mass in mammograms (image-level label) but no specific location (lesion-level label), which is a growing problem in the field of medical imaging. Thus, detection approaches in case of the lack of labels need to be explored instead of a high-cost method such as collecting more data with high-quality annotations.

As mentioned in related works, the CAM technique is a better solution to the problem we encounter compared with transfer learning and multi-instance learning. However, the existing methods such as Liang's proposal cannot solve the imprecise localization problem due to the lack of information when using image-level labels. Because the self-training strategy is based on the possible correct detection results to re-train the model, which can not introduce extra information about localization, especially for the objects that are occluded. Some fully-supervised learning studies utilize multi-view images to extract the extra information[42,45]. Inspired by these works, this is the first work that introduces multi-view enhancing into the weakly supervised detection of breast masses.

A standard mammography examination for each breast consists of the CC view and the MLO view as shown in Figure 3.1 (a). According to the direction of projection as illustrated in Figure 3.1 (b), the mass marked by the green bounding box is projected in the same position on the red line from the nipple (purple line) to the chest wall (yellow line) in ipsilateral pairs (CC view and MLO view mammograms on the same side) both views of mammograms. We believe column-wise correspondence in the above geometric relationship is easier to extract and more effective than region-level correspondence that Liu et al. tried to find in their work.

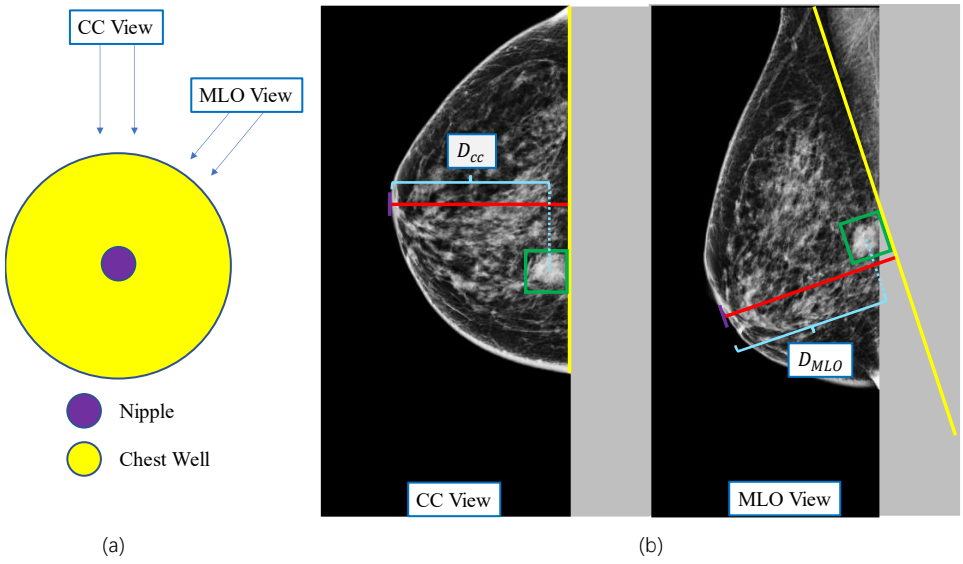

Figure 3.1 Illustrations of the projection angels and geometric relationship of multi-view.

3.2 Weakly Supervised Mass Detection Network Based on Multiview Enhancing

In this chapter, we propose a multi-view enhancing mass detection network named MVMDNet based on the weakly supervised learning to utilize the above observation about the multi-view correspondence. In MVMDNet, the Multi-view Enhancing module consisted of Spatial Correlation Attention (SCA) and the Sigmoid Weighted Fusion (SWF) address the challenge that localization information is insufficient. SCA module can extract the localization information between the feature maps of ipsilateral pairs while SWF module can fuse this correspondence. CAM-based detection module provides precise bounding boxes for breast mass. Experiments results on both in-house and public datasets show MVMDNet achieves a robust generalization among weakly supervised methods and state-of-the-art. According to the comparison with fully supervised methods, the performance of MVMDNet can provide the precise detection for breast mass same as them.

3.2.1 Overview

In this chapter, we aim to develop a weakly supervised network trained only with the image-level labels y, where $y \in \{0, 1\}$ indicates whether the mammogram contains a breast mass. MVMDNet is proposed to meet the above requirements to realize the detection of breast masses. The framework of MVMDNet is illustrated in Figure 3.2. Firstly, the ipsilateral pairs of mammograms of a single breast are preprocessed by cropping, alignment, and histogram specification. Then, both two-view mammograms are input to the feature extraction backbone

to generate the feature maps. Thirdly, Multi-view Enhancing modules generate the auxiliary feature map containing extra localization information by SCA and fuse auxiliary and diagnostic feature maps by SWF using the learnable Sigmoid weighted to generate the examined features. Binary classification is conducted on the examined features map in the training stage. Finally, mass detection on the CC view is provided by the CAM-based detection module in the inference stage.

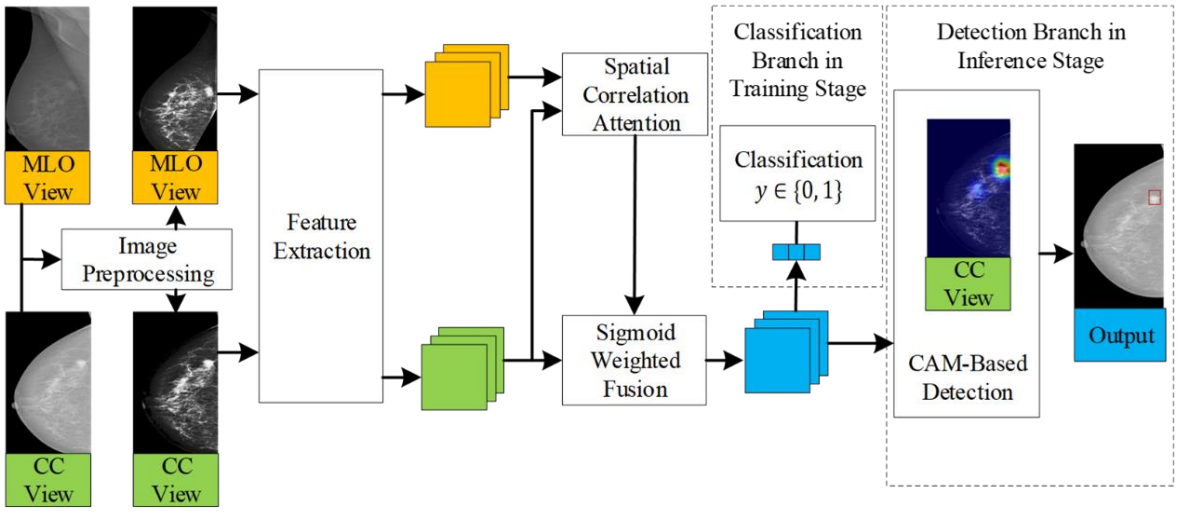

Figure 3.2 Framework of the proposed MVMDNet.

3.2.2 Preprocessing

Compared to random grayscale transformation and rotation in other related work, our preprocessing method aligns the localization information using the following three steps, which are more robust than traditional image augmentation.

First, we remove the background and the pectoral muscle for the preparation of alignment. The threshold-based method is used to segment the region of the breast from the background while morphologic operations can smooth out the boundary of the breast and guarantee the integrity of segmentation. We extract the minimum circumscribed rectangle region of the breast and employ the Hough Transform-based method to detect the pectoral muscle line and remove the pectoral muscle[62].

Second, different mammography projection views can give additional information. It's difficult to discover the region-level correspondences of CC and MLO views because of deformations and occlusions of breast tissue. In both projection views, the distance between the nipple and the chest wall is almost the same. Based on this fact, we assume that the

information of lesions provided by the CC view and the MLO view in a column of the breast is the same. The line perpendicular to the nipple to the chest wall is the direction of the breast column. We apply a column alignment to acquire the aligned MLO view pictures. The border after eliminating the pectoral muscle is considered the chest wall which is parallel to the breast column. Therefore, the image's breast region is rotated to make the pectoral muscle line vertical, and the new breast region is cropped following rotation.

Third, we adopt the histogram specification method proposed in chapter 2 to mitigate the impact of multicenter biases. The whole preprocessing is shown in Figure 3.3.

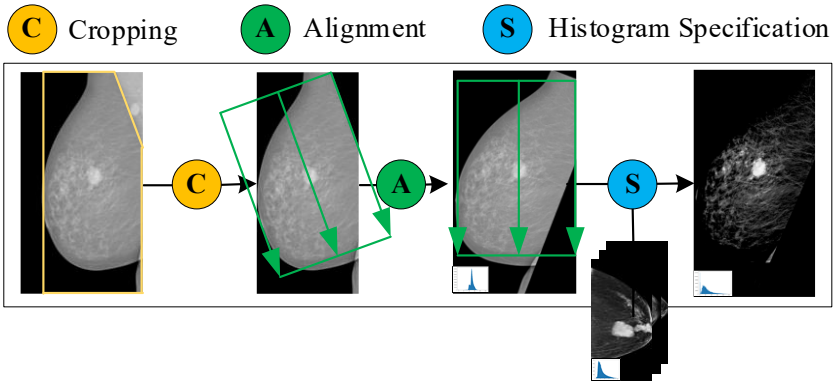

Figure 3.3 Illustration of the key steps in image preprocessing.

3.2.3 Feature Extraction

We propose a feature extraction backbone based on the EfficientNetV2 for feature extraction[63]. Table 3.1 shows components of the backbone including the Fused-MBConv and MBConv with Squeeze-and-Excitation block. Due to the consideration of dual input, we adopt a lightweight design for the backbone. Extracted features of CC view and MLO view are defined as X_{CC} and X_{MLO}, respectively, where $X \in \mathbb{R}^{C \times HW}$.

Table 3.1 The architecture of our proposed feature extraction backbone.

Stage	Block	Expand Ratio	Channels	Layers
1	Fused-MBConv	1	24	1
2	Fused-MBConv	4	48	1
3	Fused-MBConv	4	64	2
4	MBConv	4	128	2
5	MBConv	6	160	3
6	MBConv	6	272	7

3.2.4 Multiview Enhancing Module

As previously observed, localization information of lesion in the line perpendicular to the nipple to the chest wall is the same in two-view mammograms of the single breast. A set of multi-view enhancing modules is proposed to extract the correspondence in each column of the ipsilateral pairs. In our methods, the main diagnostic is conducted on the CC view while the MLO view is auxiliary.

Multiview enhancing modules consist of two parts: Spatial Correlation Attention (SCA) extracts the correspondence between CC view and MLO view in the breast column and generate the auxiliary feature defined as X_{Aux}; Sigmoid Weighted Fusion (SWF) generates an examined feature, which contains the precise localization information of breast mass, by fusing the feature map of diagnostic CC view and the auxiliary feature map of MLO view.

a) Spatial Correlation Attention

SCA consists of Spatial Column Maxpooling and Extending module (SCME) and Correlation Attention (CA) module as shown in Figure 3.4. Correspondences between ipsilateral pairs of mammograms are compressed by the former and enhanced by the latter.

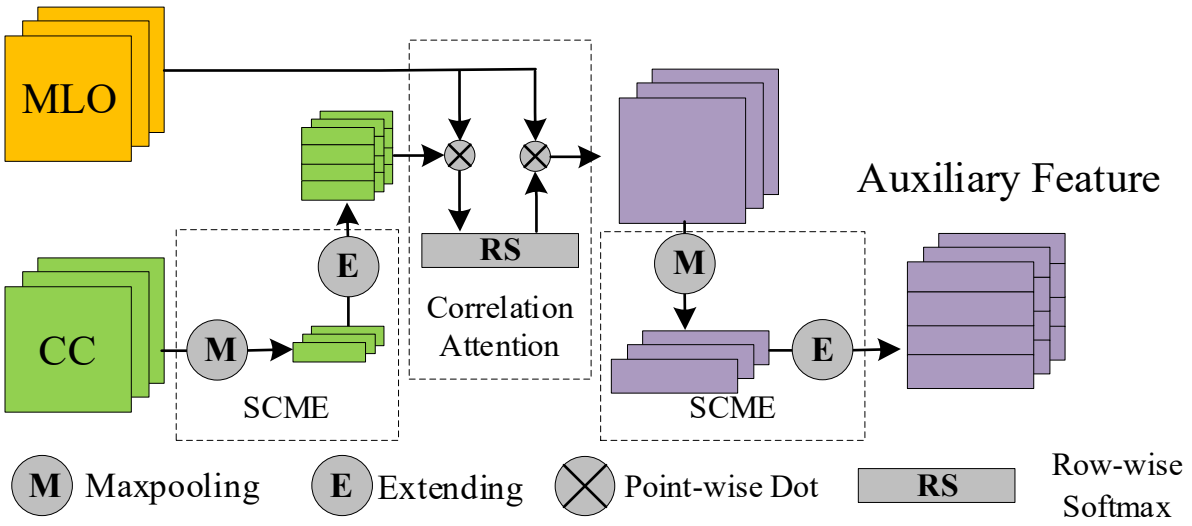

Figure 3.4 The architecture of the Spatial Correlation Attention module.

After preprocessing, breast mass is on the same column axis of ipsilateral pairs. A Maxooling operation is adopted to compress the localization information on the line perpendicular to the

line which is from the nipple to the chest wall, where $MaxPooling(X) \in \mathbb{R}^{C\times 1\times W}$, preserving the correspondences on the column axis. Extending operation, which repeats the one dimension vector of each channel, restores the same scale before Maxpooling. We define these two steps as Spatial Column Maxpooling and Extending (SCME) and firstly apply it on the CC view to compress the localization information, $\tilde{X}_{cc} = f_{SCME}(X_{cc})_i$.

The CA module is designed inspired by correlation in signal processing, which highlights the most relevant column between the MLO view feature map and features after SCME. The CA module, as opposed to the convolutional-based technique, is more interpretable when it comes to obtaining the combined feature. A point-wise dot product on each of the corresponding channels $X_i \in \mathbb{R}^{HW}$ in feature maps, where $i \in (0, C]$, is applied to find the most relevant region. Each channel of correlation matrix $R \in \mathbb{R}^{C\times HW}$ of extracted main diagnosis view features $\tilde{X}_{cc}$ and auxiliary view features X_{MLO} can obtain by Formula 3.1.

$$R_i = Multiply((\tilde{X}_{CC})_i), (X_{MLO})_i), i \in (0, C] \tag{3.1}$$

A non-normalized correlation matrix easily introduces instability. Therefore, to obtain a correlation distribution score matrix $S \in \mathbb{R}^{C\times HW}$, a Softmax operation is imposed on each row $j \in (0, H]$ of each channel $i \in (0, C]$, $R_{ij} \in \mathbb{R}^{W}$, flowing Formula 3.2.

$$S_{ij} = Softmax\ (R_{ij}) \tag{3.2}$$

We directly multiply the score matrix considered as an attention map on the feature map of the MLO view to enhance the attention on localization of mass. The enhanced features are obtained following the Formula 3.3,

$$X_{Enhanced} = Multiply(X_{MLO}, S) \tag{3.3}$$

SCME is applied on $X_{Enhanced}$ again to generate the auxiliary feature following the Formula 3.4 due to the addition of X_{MLO}.

$$X_{Aux} = f_{SCME}(X_{Enhanced}) \tag{3.4}$$

b) Sigmoid Weighted Fusion

The SWF module is proposed to fuse the main diagnosis view features X_{CC} and auxiliary feature X_{Aux}. The weights, which represent the importance, of the two feature maps are different as mentioned in chapter 2. We define the whole features Z as the concatenation of

X_{CC} and X_{Aux}. Instead of simply using two trainable parameters like MCDNet, we aim to presents the weights as relationships between partial feature, X_{CC} and X_{Aux}, and the whole features Z. First, the relationship is obtained by the convolution with Sigmoid applied on Z as shown in Figure 3.5, where Sigmoid is used to normalize the weights. Then, the weighted feature $\bar{X}_{cc}$ and $\bar{X}_{AUX}$ is obtained by the point-wise dot product. Formula 3.5-3.7 demonstrates the process of weights acquisition and application.

$$Z = Cat(X_{CC}, X_{AUX}) \tag{3.5}$$

$$\bar{X}_{cc} = Sigmoid\big(Conv(Z)\big) \cdot X_{CC} \tag{3.6}$$

$$\bar{X}_{Aux} = Sigmoid\big(Conv(Z)\big) \cdot X_{Aux} \tag{3.7}$$

Concat and Conv are applied to fuse the weighted feature $\bar{X}_{cc}$ and $\bar{X}_{AUX}$ to generate the examined features following Formula 3.8.

$$X_e = Conv(Cat(\bar{X}_{cc}, \bar{X}_{Aux})) \tag{3.8}$$

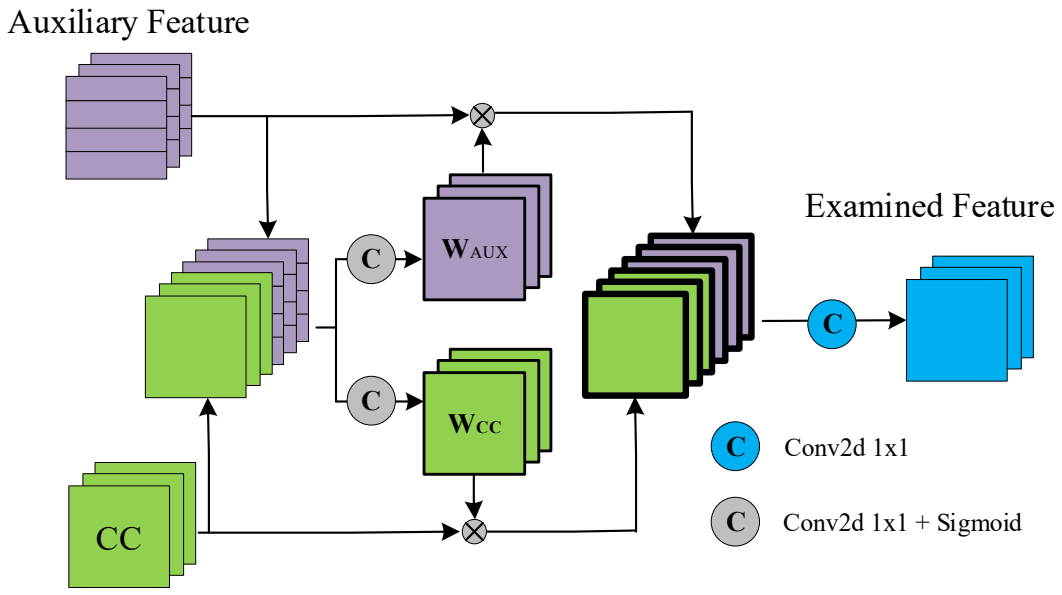

Figure 3.5 The architecture of Sigmoid Weighted Fusion module.

3.2.5 CAM-Based Detection Module

Two branches follow the examined feature map as shown in Figure 3.2. One is the classification branch that optimized by the loss between prediction $\hat{y}$ the image-level labels y. Another is the detection branch including the CAM-Based Detection module which provides the bounding boxes to realize the detections of breast masses in the inference stage. As

mentioned before, CAM is a technique for visualization of the feature map, which can determine the attention regions of the network. Generally, these attention regions are the basis for network classification.

The steps to achieve by CAM are as follows: Firstly, the CAM saliency map is generated on the specified feature map through different techniques, such as gradient-based CAM. Secondly, the CAM saliency map is scaled to a certain range, which is 0 to 255 in this thesis. Finally, a bounding box for the threshold-based segmentation of attention regions is generated for the localization of the object. Eigen-CAM is employed in CancerDNet because of its great performance[64].

3.3 Experiments

3.3.1 Datasets and Implementation

Considering the network inputs, we collect ipsilateral pairs, mammograms of CC view and MLO view, from the in-house and INbreast datasets to form the datasets used in this chapter. 1460 ipsilateral pairs mammograms with the clinical diagnoses, which indicate whether there are masses in mammograms, are collected from the in-house dataset. The ratio of training, verification, and testing are 8:1:1. The ground truths of detection are labeled by radiologists. 187 ipsilateral pairs are collected from INbreast dataset. The original masks contained in the dataset are considered as the training labels and ground truth. Model pre-trained on the in-house dataset is fine-tuned on INbreast dataset. The experiments on INbreast dataset use a 5-fold cross-validation strategy.

Like chapter 2, the in-house dataset is the target mammograms while INbreast is the source mammograms in Histogram Specification. Input resolution of the network is 370×840. The threshold of the CAM-Based Detection module is 50.

Models are implemented by PyTorch and trained on a platform with an NVIDIA Tesla V100 SXM graphics processing unit (GPU) that is used in chapter 2. The initial learning rate is 0.1 with a weight decay of 10^{-4}. The batch size is 32 and the max number of epochs is 100. Cross-entropy loss is employed in the training stage.

3.3.2 Evaluation Metrics

Correct localization (CorLoc) is defined as the percentage of targets correctly localized, which is a commonly used evaluation metric in weakly supervised studies[40]. In most studies, a mass is considered localized correctly if the IOU between prediction and the ground truth bounding box is above 0.2[65,69]. CorLoc is employed in the ablation study to evaluate the contributions of proposed methods for the weakly supervised network. The other metric is Recall@FPI used in chapter 2, which is commonly used in the studies of detection in mammography whether supervised or weakly supervised methods are adopted[67,70].

3.3.3 Experiment Plan

Firstly, the experiment is conducted on the in-house dataset to test the mass detection performance of MVMDNet. We reproduced some high-performance methods to compared: YOLOv5x is a popular detection network recently[71]. In this experiment, YOLOv5x is trained with different proportions of the dataset. GMIC is a weakly supervised localization network improved from the previous version and we convert the multi-classification to binary classification[72,39].

Secondly, MVMDNet is compared with some state-of-the-art methods experimented on INbreast dataset. Dhungel proposed a fully-supervised automatic mass detection network using a cascade of deep learning and random forest classifiers[66]. The two-stage design aims to further reduce the false positive detections. Agarwal designed a fully automated framework to detect masses using the Faster-RCNN models[37]. Then, the F-RCNN model that trained on their in-house dataset used transfer learning to detect masses in the INbreast dataset. Tardy proposed a weakly-supervised method trained on image-level labels[70]. They trained a reconstruction network using normal data to generate breast region without abnormal and perform detection relying on the difference between the origin abnormal image and image reconstructed from the former. In addition, we also conduct a GMIC model on the INbreast dataset for comparison.

Finally, we analyzed each component of the methods we proposed. We compare the results of different CAM techniques including Grad-CAM, Score-CAM, and Eigen-CAM to determine the localization ability[73-74]. We applied them on our backbone respectively and conduct training on the in-house dataset. Then, ablation studies were performed to evaluate the contribution of the SCA module and SWF module.

3.3.4 Result and Discussion

a) Comparison on In-house Dataset

The ratios (10%, 50%, 100%) in Table 3.2 indicate the proportions of the training set we used to train YOLOv5x. The experimental results demonstrate MVMDNet achieves the state-of-the-art performance with Recall@FPI of 0.92@0.52. The results of YOLOv5x get improvement with the increase of labeled data ratio from a recall rate of 0.61 to 0.76, which demonstrates the importance of a large number of annotations in the supervised method. Even trained with 100% labeled data, YOLOv5x does not perform well on our in-house dataset. GMIC performs a little better than YOLOv5x trained by a 10% labeled training set but with a higher false positive.

Table 3.2 Performance on In-house dataset.

Methods	Supervision Mode	Recall@FPI
YOLOv5x (10%labeled)	Full	0.61@0.38
YOLOv5x (50%labeled)	Full	0.72@0.47
YOLOv5x (100%labeled)	Full	0.76@0.99
GMIC	Weak	0.63@0.70
MVMDNet	Weak	0.92@0.52

As shown in Figure 3.6, we provide some detection examples from our in-house dataset to explain the high performance of MVMDNet demonstrated by the above results. Each row represents an example of a case. (a) and (b) are images of CC and MLO view, where green boxes indicate the masses. (c)-(e) are detection results of YOLOv5, GMIC, and MVMDNet, respectively. The example in the first row contains a mass with a clear border, which is easy to be distinguished from breast tissue. The detection bounding box for this case generated by MVMDNet is the same as YOLOv5x. However, GMIC does not localize the mass correctly in the same case. One reason is that the non-end-to-end patch selection algorithm in GMIC does not catch the precise location of the target in the training stage, which results in a high false positive rate. The second example is one of the ‘hard samples’ that contain masses that are almost invisible in the CC view. Because these masses are occluded by healthy breast tissue, the information of one view features after extraction is insufficient for localization. MVMDNet localizes the masses precisely while both YOLOv5x and GMIC miss these

targets. According to the observation of the second example, this mass is visible in MLO view images as shown in (b) and the mass projected position on the row axis is the same as the CC View. This correspondence helps MVMDNet to find the correct location of mass in CC view and further improve the performance. This experiment demonstrates that information enhanced by multi-view training can provide more effective information than multi-instance learning in the absence of lesion labels.

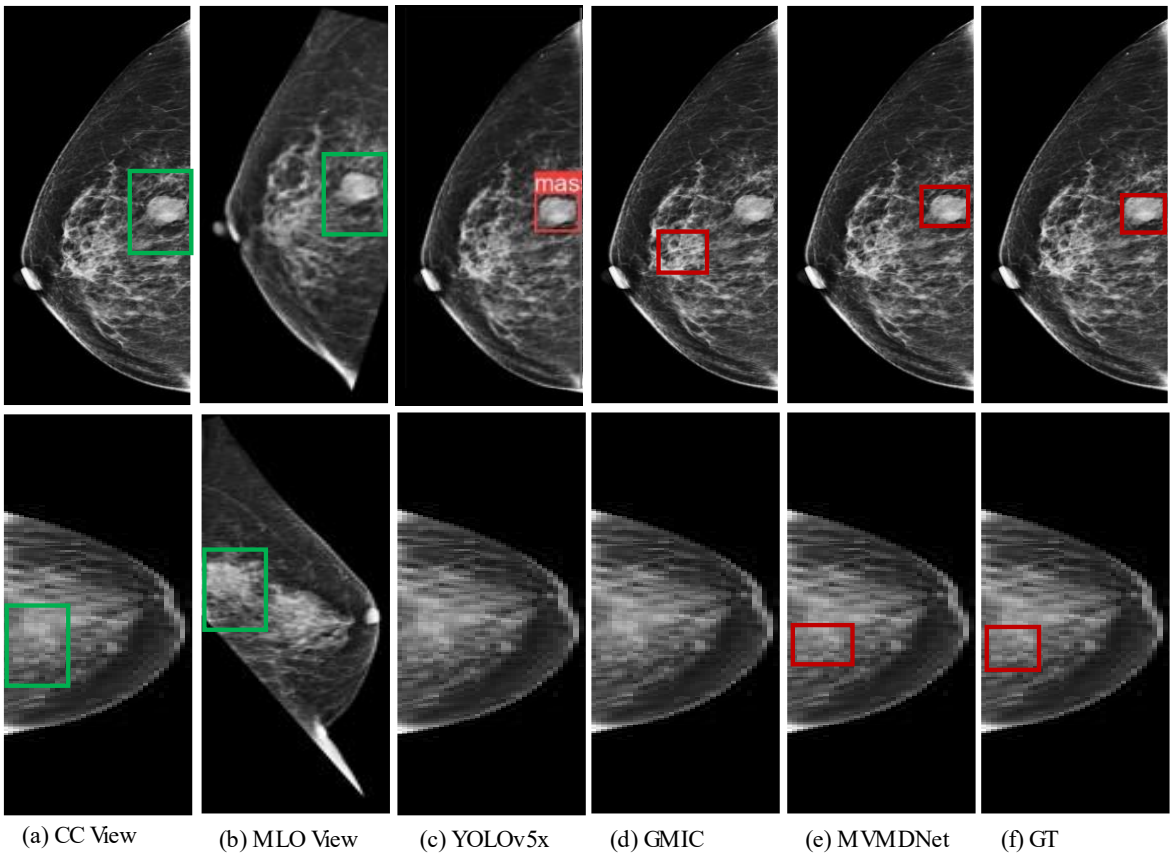

Figure 3.6 Examples of detection results. Green bounding boxes indicate the location of masses.

b) Comparison with State-of-the-Art Methods

The mass detection results on the INbreast dataset are shown in Table 3.3. MVMDNet achieved a Recall@FPI of 0.96±0.02@0.77 with a lower FPR than others. We also train GMIC with only image-level labels and catch the patch from saliency maps as detection results and MVMDNet performs better than it in both recall rate and false positive rate. The recall rate of our model is the same as Tardy's methods, a weakly supervised method based on reconstruction, while false positive rate is lower than the latter. These results show our model is better than other weakly supervised methods. The fully supervised method proposed by Agarwal achieves the best recall rate of 0.99, which makes sense because of the training on

lesion-level labels. However, their method introduces a high false positive rate (FPR). One possible reason is that they chose a lower IOU threshold of 0.1 than other methods, which is more inclined to recall ability. The caparison with supervised methods proves that MVMDNet gets close to the supervised methods. In addition, because MVMDNet is trained on our in-house dataset and performed fine-tuning on the public INbreast dataset, the results also indicate that our proposals have robust generalization ability.

Table 3.3 Performance on public INbreast dataset.

Methods	Supervision Mode	Recall@FPI
MVMDNet	Weak	0.96±0.02@0.77
GMIC	Weak	0.92±0.02@1.44
Tardy's	Weak	0.96±0.01@0.85
Dhungel's	Full	0.90±0.02@1.30
Agarwal's	Full	0.99±0.03@1.17

c) Ablation study

The results of several CAM techniques applied on the backbone are shown in Table 3.4. Both Eigen-CAM and Score-CAM achieve a high detection CorLoc of 0.725, but Score-CAM takes about 20 times longer than Eigen-CAM. In the example shown in Figure 3.7, Eigen-CAM provides a bounding box that is more precise than bounding boxes generated by Score-CAM and Grad-CAM.

Table 3.4 Localization performance of different CAM techniques applied on In-house dataset.

CAM Techniques	CorLoc	Time per Image (s)
Eigen-CAM	0.725	0.43
Score-CAM	0.725	10.8
Grad-CAM	0.648	0.48

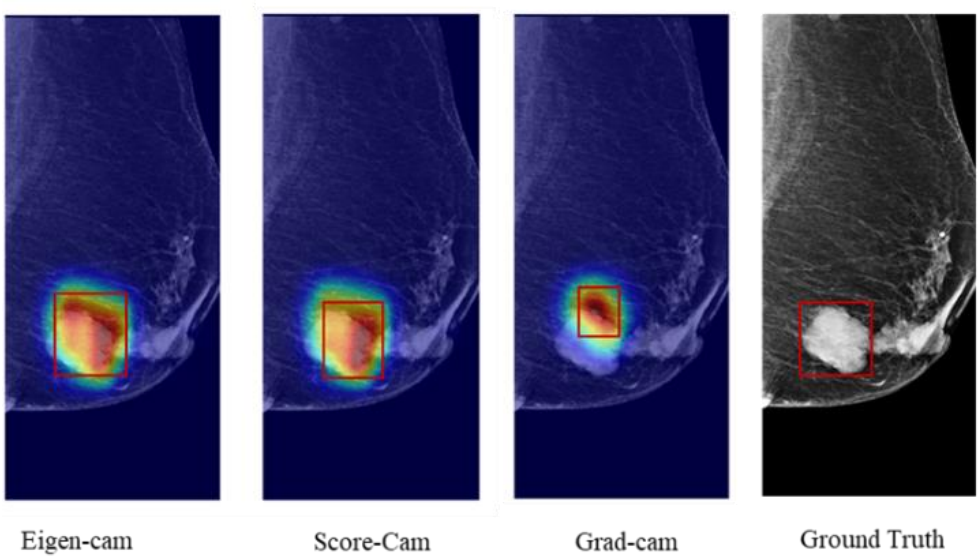

Figure 3.7 An example of the localization ability of different CAM techniques.

In the ablation study, we consider the feature extraction backbone applied with only CAM-Based Detection module as MDNet. We firstly focus on SCA module and replace SWF module with a set of Concat and Conv (Cat & Conv) to eliminate the effect of SWF. When SCA module is applied on MDNet, the results shown in Table 3.5 demonstrate a great improvement of CroLoc to 0.698. Analyzing the cases that are classified correctly after adding the SCA module, we find the masses in these cases are hard to be distinguished from breast tissue like the second example shown in Figure 3.6. This observation indicates the SCA module can extract extra mass localization information from MLO view, which is hard to be extracted from CC view, to enhance the network's attention to the correct location of mass in the image. Then, we adopt the SWF module and the performance improves by reducing some wrong cases, where the mass is only present in MLO view images. The fusion strategy learned by the network itself is better than simply designing a general strategy. The best model of MVMDNet with all components achieves CroLoc of 0.725.

Table 3.5 Effectiveness of each component on In-house dataset.

Methods	AUC	CorLoc
MDNet	0.730	0.591
MDNet + SCA + Cat & Conv	0.836	0.698
MDNet + SCME+ Cat & Conv + SWF	0.830	0.680
MVMDNet	0.858	0.725

Especially, we replace the Correlation Attention module with a set of Concat and Conv (Cat & Conv) in the original SCA (SCME + Correlation Attention) to determine the contribution of the Correlation Attention module. The performance of the new combination reduces a lot from the CorLoc of 0.725 to 0.680. It proves that our design of the correlation attention module according to prior knowledge is better than learning itself by the Conv. It seems to be in contradiction with the previous that weights learned automatically result in a good performance, but it shows that we should make a trade-off between deep learning and human design according to the specific situation.

We also provide a visualization of how the multi-view enhancing modules work as shown in Figure 3.8. (a) and (c) are images of CC and MLO view. (b) and (e) are saliency maps of MDNet and MVMDNet, respectively. Comparing (b) and (e), it is obvious that the presents of localization information are on the same row axis positions of ipsilateral pairs while uncertain is on the column axis. The auxiliary features after extending in Spatial Correlation Attention are shown in (d) and it shows that localization information in the column is successfully preserved in auxiliary features. The saliency maps in (b) and (e) show the attention of the network without and with the SCA module, respectively. We can observe the effectiveness of the impact by the multi-view enhancing module. It is like applying the dot product on the auxiliary feature map to the CC view feature map. The multi-view enhancing modules effectively restrict the wrong localization information as shown in the example of the first row and provide correct localization information on CC view as shown in the example of the second row.

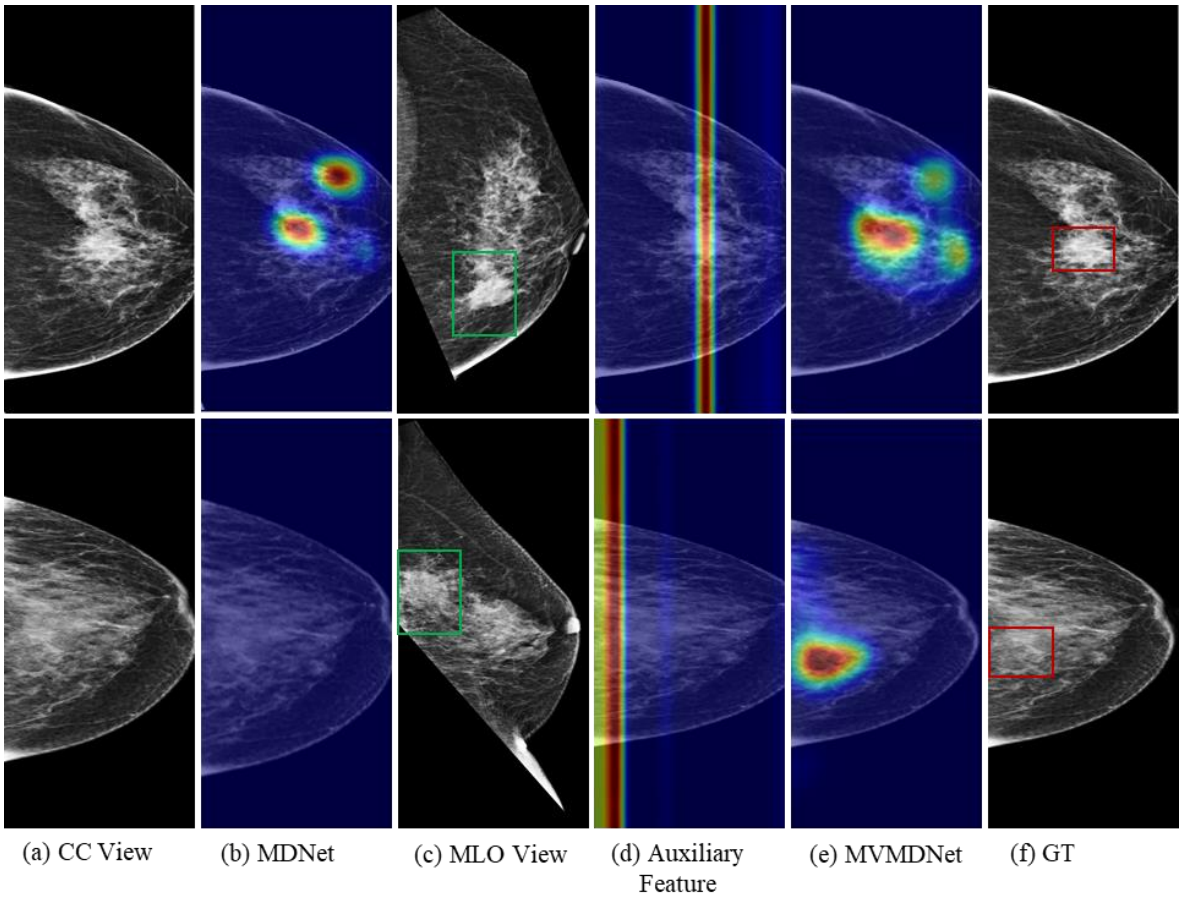

Figure 3.8 Visualization of the effectiveness of multi-view enhancing modules. (b), (d) and (e) have been mixed with saliency maps.

3.4 Conclusion

In this chapter, a weakly-supervised network MVMDNet is proposed to realize the detection of mass. MVMDNet can be trained on the dataset that only contains lesion-level labels on mammograms, and successfully solve the problem of insufficient localization information faced by the weakly supervised detection network. MVMDNet aligns the ipsilateral pairs and reduces the bias from multicenter datasets by image preprocessing. The Multi-view Enhancing module extracts the correspondence of ipsilateral pairs and enhances the localization feature by a fusion of diagnostic and auxiliary features. The CAM-based detection module converts the feature map optimized by the classification task to provide precise localization for breast masses. On both in-house and INbreast datasets, results show that MVMDNet achieves state-of-the-art performance when compared to other weakly-supervised methods and has the same accurate localization capabilities as supervised methods. The results of the fine-tuned model conducted on INbreast further show that MVMDNet has strong generalization.

In the works of the above two chapters, the detections of common lesions in breasts are completed. In the next chapters, we will explore the cancer diagnosis based on the detection works.

4 Chapter 4: Breast Cancer Diagnosis in Mammography Based on Multi-instance Learning

4.1 Introduction

The goal of the breast cancer diagnosis in mammography is to provide radiologists an indication of the presence or absence of cancer in a set of mammography images for a case. Transforming into deep learning-related problems, the cancer diagnosis is a classification task. The idea of this thesis is to develop a binary classification network for the whole breast mammograms of one case, which is a more comprehensive approach compared with the methods that focus only on binary classification of macrocalcification or mass in previous related work mentioned in chapter 1[47,49].

However, the features of lesions that indicate the presence of cancer are complicated, which introduces a low performance if directly using the whole images for classification. For example, the feature extreme tiny microcalcifications are easy to be ignored, especially when the network compresses the resolution of images limited by computing resources. Recently, some works mentioned in chapter 1 try to classify the whole image with lesion information and achieve impressive results. Shanms et al. utilize the GAN network trained on lesion patch to optimize the feature extraction of the whole image classification network[54]. McKinney et al provide an AI system that consists of three deep learning networks for lesion level, breast level, and case level classification respectively, which diagnose according to the average score of three networks[59]. Benefitting from the works in previous chapters, we propose a deep learning approach for cancer diagnosis based on whole image classification, which drops the common method that calculates the average score of different network outputs.

4.2 Multi-instance Learning Network CancerDNet for Cancer Diagnosis

In this chapter, we proposed a multi-instance learning network for breast cancer diagnosis named CancerDNet to solve the problem of whole image classification, which realizes the precise classification for the whole case to help radiologists diagnose cancer in mammograms. In previous chapters, the trained models including MCDNet and MVMDNet could provide precision detection results or classification vectors for the cancer diagnosis network that we would explore. Firstly, the idea of multi-view enhancing is continued to be used to integrate the multi-view features extracted from the designed backbone. Both two sides of breast feature maps are concatenated after feature extraction because the diagnosis is based on the

whole case. Then, we divide the detection results of the MCDNet and MVMDNet into two types considering the contribution of information and precision of results: Results of the mass detection are considered as the low-capacity lesion instances, which are used as masks to make the classification network pay more attention to the regions where breast masses exist; Results of the microcalcification are considered as the high-capacity lesion instances, of which the classification vectors are concatenated to the vectors of cancer classification to make the network.

4.2.1 Overview

In this chapter, we will describe the multi-instance learning network CancerDNet in detail. As shown in Figure 4.1, CancerDNet consists of MCDNet, MVMDNet, and a classification network. Firstly, the image preprocessing integrated from works of previous chapters adjusts the multicenter data input. Secondly, the different sets of preprocessed images are input to the different components of the multi-instance learning stage, including MCDNet, MVMDNet, and the feature extraction backbone of classification, to generate the microcalcification detection vector, mass localization, and feature map of each view. Thirdly, Whole Case Bag Learning (WCBL) fuses the four-view feature maps on which Low-capacity Instance Learning (LCIL) adds the mass localization information from MVMDNet before fusion, and then High-capacity Instance Learning (HCIL) integrates the microcalcification information from MCDNet into the classification vector generated from the fused feature map. Finally, the network provides cancer diagnosis through a binary classification that indicates whether cancer is found in the case.

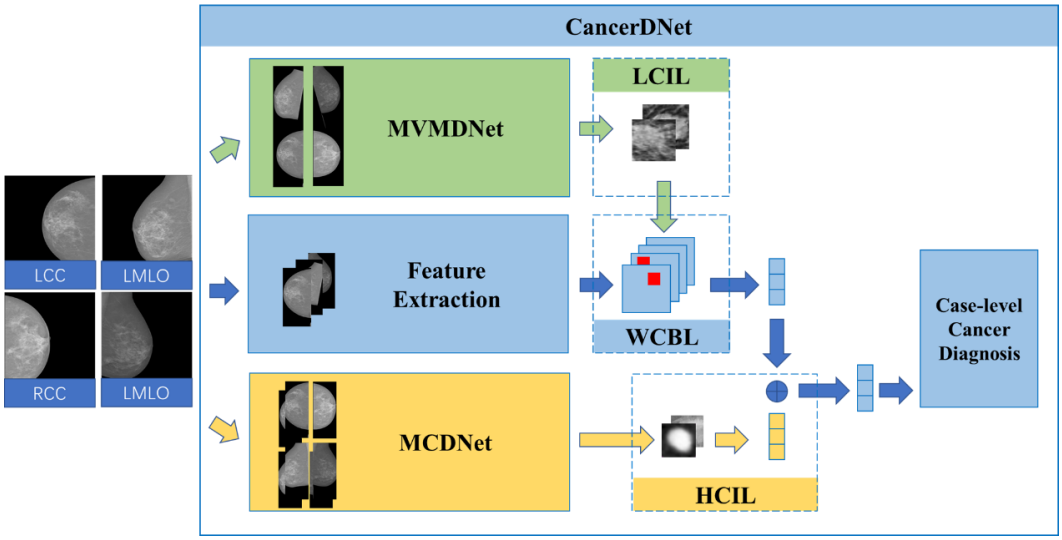

Figure 4.1 Framework of CancerDNet.

4.2.2 Whole Case Bag Learning

As shown in Figure 4.1, the input of the CancerDNet is a set of mammography images for one case including the four views, LCC, LMLO, RCC, and LMLO, which is considered as the whole case bag. Compared with the diagnosis according to single view mammography image, the models trained by the whole case will learn the correspondence between the multi-view, which is proved by the works of previous chapters. The preprocessing method for four view images before sending to the classification backbone is the same as the method proposed in chapter 3 consists of cropping, alignment, and histogram specification. This operation makes the weakly supervised detection results of the masses easy to align with the feature extracted by the classification backbone.

Here, we conduct the structure named Whole Case Bag Learning (WCBL) shown in Figure 4.2 to realize the whole case bag learning. Firstly, the backbone we used to extract features has the same structure as MVMDNet, which is finetuned on the trained MVMDNet models and shared weights for all four views images. This design can reduce training and inference time consumption. Then, the extracted features for mammography images of one breast are concatenated to expand the channel. One Conv that keeps the size of features squeezes the number of channels to find the correspondences between ipsilateral pairs. Unlike the Spatial Correlation Attention employed in MVMDNet, we adopt a lighter structure here to pay more attention to optimizing consumption efficiency. Thirdly, the Sigmoid Weighted Fusion we proposed in chapter 3 is employed to fusion the feature maps for each breast produced by former concatenation and convolution. This operation works like a radiologist: they compare the symmetrical structure of the two breasts and derive a final score based on both sides.

4.2.3 Low-capacity Instance Learning

Unlike the AI system proposed by McKinney, we believe that the detection results from MVMDNet contain insufficient information. Because the results of the MVMDNet only indicate the location of the masses in the breast rather than the classification of masses. Therefore, we do not directly convert the detection result to a classification result. Like a team of radiologists, the classification networks should pay more attention to the regions where other models consider lesions present. Our idea is to convert the localization of mass to a mask, which is presented by 1 and 0. As shown in Figure 4.2, concatenation and convolution are used to fuse the mask and the corresponding mammograms instead of using the dot

product because the former is more stable than directly using the mask to remove the non-attention region.

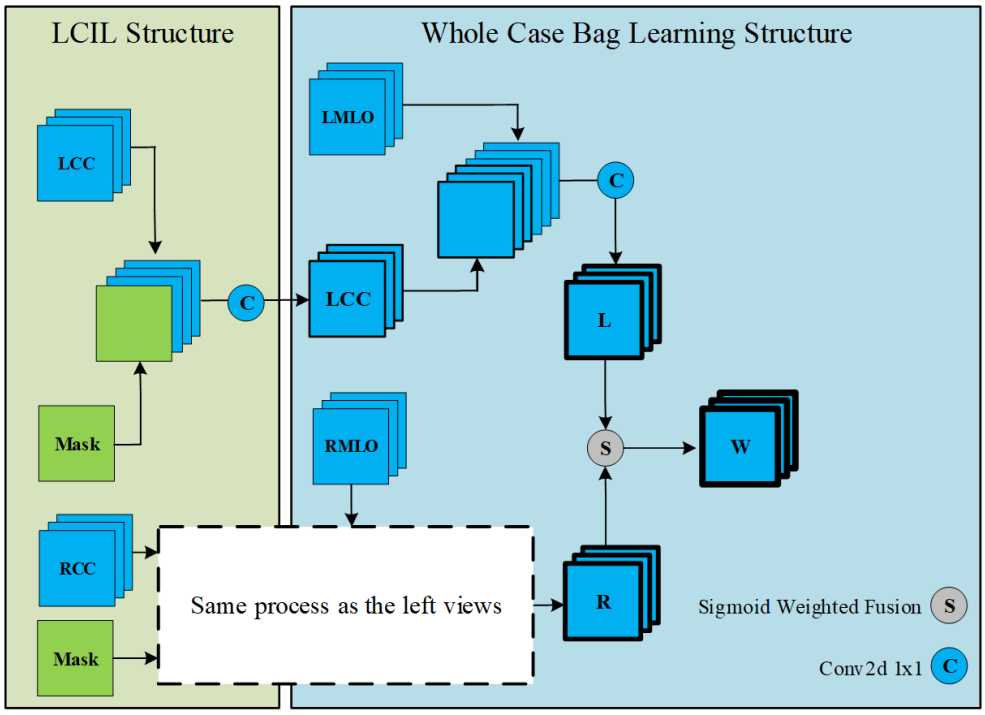

Figure 4.2 Structures of LCIL and WCBL.

4.2.4 High-capacity Instance Learning

The results of MCDNet include more precise information of localization and classification, which we consider as the high-capacity lesion instance. Thus, the detection results of a standard set of mammography screening images could be directly involved in the whole case bag classification of the final cancer diagnosis in CancerDNet.

The demonstration of HCIL is shown in Figure 4.3. Because the types and quantities of microcalcifications detected in the standard set of four images are different, the ranking mechanism is employed to select the results of detection. Considering the average number of the malignant microcalcification detections, two detections with the highest scores among all four mammography images are selected. Firstly, an average pooling is applied on the classification vectors of these two detections are extracted from the classification branch of MCDNet and are reduced to the same length as the vector in CancerDNet classification backbone by linear operation. An average pooling is conducted on these two vectors to generate the high-capacity lesion instance classification vector. Then, this vector is concatenated on the classification vector generated from the fully connected layer of the

backbone. Finally, the new vector is sent to an MLP to generate the classification result. The size of the patch shown in Figure 4.3 demonstrates the rank of scores.

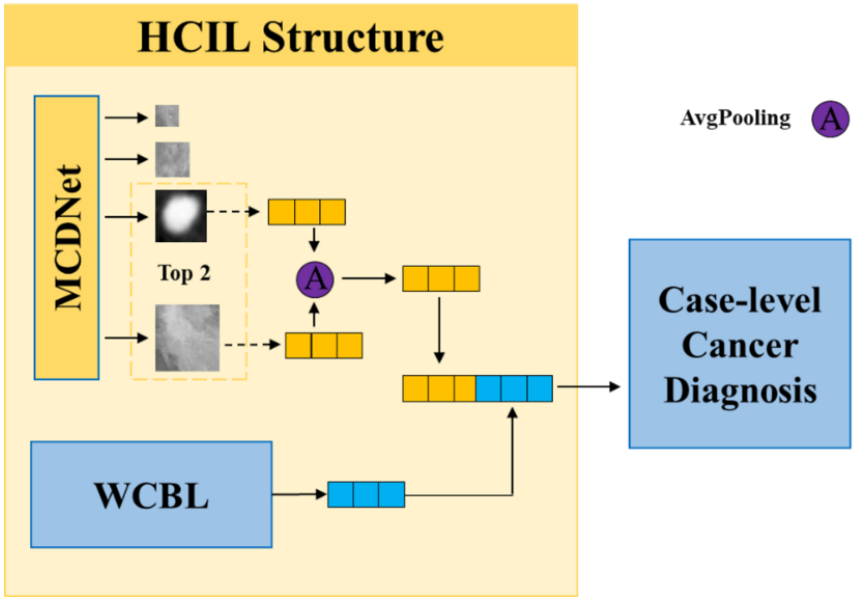

Figure 4.3 Demonstration of HCIL. The sizes of lesion patches indicate the scores of detections.

4.3 Experiments

4.3.1 Datasets and Implementation

We conduct the cancer diagnosis experiments both on the in-house dataset and the public dataset INbreast. We collect 688 cases with the full set of standard mammography images, a total of 2752 images from in-house datasets. The origin clinical diagnoses from biopsy results are regarded as the ground truth. The ratio of training, verification, and testing set are 8:1:1. Due to the lack of ground truth like our in-house dataset in the reports of INbreast, we assigned all images with BI-RADS 4, 5, and 6 as positive while all images with BI-RADS 1, 2 as negative. Because BI-RADS 3 means probably benign without a definitive diagnosis, we exclude some cases that at least have one image diagnosed as BI-RADS 3. Therefore, we finally select 82 cases with four images, a total of 328 images, for two side breasts in the experiments of this chapter. All cases collected from INbreast are used in testing.

The experiments of CancerDNet are implemented on the platform with an NVIDIA Tesla V100 SXM GPU and developed by PyTorch like the previous models MVMDNet we proposed. Gradient descent with momentum is used to optimize the model. The momentum is 0.9 and the learning rate is 0.1 with the weight decay 10^{-4}. The max number of training epochs is 100 and the best model is selected when training. The input setting for MVMDNet

and MCDNet are same as the before while the resolution input of the classification backbone is the same as MVMDNet.

4.3.2 Evaluation Metrics

In this chapter, the receiver operating characteristic curve (ROC) and the area under the curve (AUC) is used to evaluate the performance of the model in experiments. The x-axis of ROC is the false positive rate while the y-axis is the true positive rate, and each point on the ROC represents a different threshold for the binary classification. Thus, ROC indicates how the models trade off false positive rates and the true positive rate at different operating points. AUC is to evaluate the classification performance of the model quantitatively by calculating the area under the roc and the higher AUC indicates the better classification performance.

4.3.3 Results and Discussion

a) Ablation Studies

In this section, we conduct the ablation studies on the in-house dataset to determine the effectiveness of the architectures we proposed for the multiple lesions.

Compared to the CancerDNet, the baseline mentioned in Table 4.1 replaces the WCBL with a direct concatenation of the four-view image and demolishes LCIL and HCIL. According to the results shown in Table 4.1 and the ROC curves shown in Figure 4.4, each module we proposed contributes to the improvement of the performance of the cancer diagnosis. As shown in Table 4.1, the application of WCBL improves the AUC from 0.764 to 0.795, which is used in subsequent ablation studies by default.

Table 4.1 Results of ablation study.

Methods	In-house dataset (AUC)
Baseline	0.764
Baseline + WCBL	0.795
Baseline + WCBL + LCIL	0.835
Baseline + WCBL + HCIL	0.883
CancerDNet	0.907

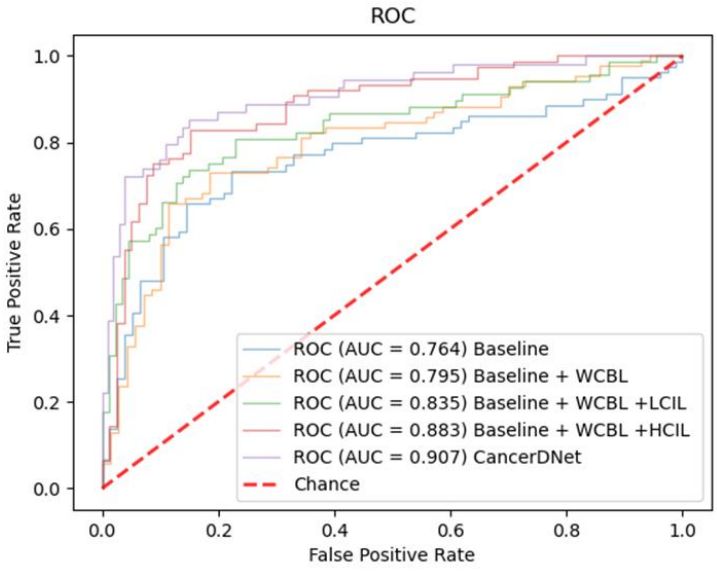

Figure 4.4 ROC curves of ablation study.

By examining the classification results of different methods, we could analyze the relationship between cases that are correctly classified and the methods applied in the network. Some masses that are indistinguishable from the background tissue get the attention of the network with the application of LCIL, which connects the MVMDNet and classification backbone. Therefore, the AUC increased from 0.795 to 0.835. A huge improvement of AUC, an increase of 0.088 from 0.795 to 0.883, occurs in the application of the HCIL, which connects the MCDNet and classification backbone. Due to the consideration for the efficiency of models, the backbone of CancerDNet implements the same setting as the MVMDNet, which has the strong ability to extract the features of a normal size object rather than an extremely tiny object such as microcalcification. Thus, the classification vectors of HCIL help the CancerDNet improve the ability to consider the presence of the extreme tiny microcalcification when classifying the whole case from the final stage of classification.

b) Compare with Other Methods

To verify the performance of the CancerDNet proposed in this chapter for cancer diagnosis, a comparison with other whole image classification methods is performed in this chapter. Table 4.2 demonstrates the results of the AI system proposed by McKinney et al. and CancerDNet.

As shown in Table 4.2, the results of CancerDNet achieve the state-of-the-art performance in the whole case classification in public INbreast datasets and are better than the AI system proposed by McKinney et al. conducting on our in-house dataset. The improvement of

CancerDNet compared to the AI system is reflected in some cases in which averaging the three models' risk scores resulted in a reduction of sensitivity. This experiment proves that using a deep learning model to integrate the multi-instance learning is a better choice than simply calculating the scores to combine the multiple models.

Table 4.2 Results of comparison.

Methods	In-house dataset (AUC)	INbreast (AUC)
AI system	0.880	0.886
CancerDNet	0.907	0.925

The results on INbreast show that the CancerDNet has a stronger generalization than the AI system on multicenter datasets. According to the difference between the two datasets, the multicenter bias generally affects the performance of a deep learning model trained on a single dataset. In the experiment, the model is only trained on our in-house dataset and the whole INbreast is used to test the generalization on multicenter datasets. As shown in Table 4.2, the AUC of CancerDNet is 0.039 higher than the AI system when testing on INbreast, which is higher than the increase, 0.027, testing on our in-house dataset. We believe it benefits from the preprocess method we proposed. Thus, CancerDNet is more suitable for multicenter datasets.

4.4 Conclusion

To solve the problem that the lesion features are complex in mammography images faced by the whole image classification, a multi-instance learning network CancerDNet is proposed in this chapter based on the works and detection results of previous chapters. CancerDNet employs the Whole Case Bag Structure to realize the classification of each case, which imitates the diagnosis procedure of a radiologist. In addition, the implementation of Low-capacity Instance Learning and High-capacity Instance Learning improves the classification precision by helping the network take into account information of multiple lesions. CancerDNet can provide higher performance of cancer classification in mammography images compared with other full image classification models.

5 Chapter 5: Conclusion and Perspective

5.1 Conclusion

Breast cancer is the most diagnosed cancer in women all over the world. The use of CAD based on deep learning to assist radiologists in lesion detection and cancer diagnosis in mammography screening is effective. However, the existing CADs do not perform well because face the difficulties caused by the characteristics of mammography images. This thesis aims to utilize deep learning approaches which consider the characteristics of mammography images to realize the high-performance cancer diagnosis system based on the multiple lesions detection in mammograms, and the achievements on research are as follow:

In exploring the precision detection of microcalcification in mammograms, we conduct studies on deep learning networks based on multi-feature fusion. MCDNet is proposed to overcome the problems of multicenter data bias, the low resolution of network inputs, and scale differences between microcalcifications. AIA mitigates the impact of multicenter biases and maximizes the input effective pixels. SCPN improves the detection ability of the network for multiscale objects by enhancing the propagation of critical features; WFF is proposed to improve the detection performance of both scale objects. The experiments show that MCDNet is more precise and robust than other methods in microcalcification detection, which can effectively help radiologists identify two types of microcalcifications and provide precise localization for CancerDNet.

In exploring the detection of breast masses in mammograms, a weakly supervised multi-view enhancing mass detection network named MVMDNet is proposed in chapter 2 to solve the lack of lesion-level labels. The image preprocessing mitigates the biases and aligns the multi-view of mammograms. The Multi-view Enhancing module extracts localization information between different views and enhances this information in a fusion, which solves the lack of localization information in weakly supervised training with only image-level labels. CAM-based detection module provides detections for mass based on the enhanced localization information. The results of experiments demonstrate that MVMDNet achieves state-of-art performances and a robust generalization among weakly supervised methods.

In exploring the cancer diagnosis in mammography, a breast cancer diagnosis network based on Multi-instance Learning named CancerDNet is proposed in chapter 4. CancerDNet successfully solves the problem that the features of lesions are complex in whole image

classification, utilizing the lesion detection results from the previous chapters. WCBL combines the features extracted from four-view and works as a radiologist to realize the classification of each case. LCIL and HCIL integrate the detections of multiple types of lesions so that the model can fully consider lesions with complex features in the classification task. The experiments conducted on both in-house and public datasets show that CancerDNet achieves a high-performance cancer diagnosis and can help radiologists reduce the workload.

5.2 Perspective

Future work of this thesis would be further studied in the following two aspects：

Here, a multi-stage network is proposed to realize the cancer diagnosis where the MCDNet and MVMDNet directly provide the results. In future work, three parts would be constructed into an end-to-end multi-task learning network to further improve the performance for each task and the efficiency of optimization, and clinical reports can be automatically generated based on detection and diagnosis results.

The weakly supervised method is applied to breast mass detection, which is a tough task. In future work, the application of weakly supervised and unsupervised learning on lesions detection including microcalcification and mass would be further explored to solve the increasingly severe problem of the lack of annotations.

Reference

[1] Siegel R L, Miller K D, Fuchs H E, et al. Cancer statistics, 2022[J]. CA: a cancer journal for clinicians, 2022.

[2] Xia C, Dong X, Li H, et al. Cancer statistics in China and United States, 2022: profiles, trends, and determinants[J]. Chinese Medical Journal, 2022.

[3] Tabár L, Vitak B, Chen T H H, et al. Swedish two-county trial: impact of mammographic screening on breast cancer mortality during 3 decades[J]. Radiology, 2011, 260(3): 658-663.

[4] Marmot M G, Altman D G, Cameron D A, et al. The benefits and harms of breast cancer screening: an independent review[J]. British journal of cancer, 2013, 108(11): 2205-2240.

[5] Allemani C, Matsuda T, Di Carlo V, et al. Global surveillance of trends in cancer survival 2000–14 (CONCORD-3): analysis of individual records for 37 513 025 patients diagnosed with one of 18 cancers from 322 population-based registries in 71 countries[J]. The Lancet, 2018, 391(10125): 1023-1075.

[6] Sree S V, Ng E Y K, Acharya R U, et al. Breast imaging: a survey[J]. World journal of clinical oncology, 2011, 2(4): 171.

[7] Bevers T B, Helvie M, Bonaccio E, et al. Breast cancer screening and diagnosis, version 3.2018, NCCN clinical practice guidelines in oncology[J]. Journal of the National Comprehensive Cancer Network, 2018, 16(11): 1362-1389.

[8] Flobbe K, Bosch A M, Kessels A G H, et al. The additional diagnostic value of ultrasonography in the diagnosis of breast cancer[J]. Archives of internal medicine, 2003, 163(10): 1194-1199.

[9] Berg W A, Blume J D, Cormack J B, et al. Combined screening with ultrasound and mammography vs mammography alone in women at elevated risk of breast cancer[J]. Jama, 2008, 299(18): 2151-2163.

[10] Bahl M, Baker J A, Greenup R A, et al. Evaluation of pathologic nipple discharge: what is the added diagnostic value of MRI?[J]. Annals of surgical oncology, 2015, 22(3): 435-441.

[11] China Anti-Cancer Association. Breast cancer screening guideline for Chinese women[J]. Cancer Biology & Medicine, 2019, 16(4): 822-824.

[12] Boyd N F, Guo H, Martin L J, et al. Mammographic density and the risk and detection of breast cancer[J]. New England journal of medicine, 2007, 356(3): 227-236.

[13] Posso M C, Puig T, Quintana M J, et al. Double versus single reading of mammograms in a breast cancer screening programme: a cost-consequence analysis[J]. European radiology, 2016, 26(9): 3262-3271.

[14] Posso M, Carles M, Rué M, et al. Cost-effectiveness of double reading versus single reading of mammograms in a breast cancer screening programme[J]. PloS one, 2016, 11(7): e0159806.

[15] Balleyguier C, Kinkel K, Fermanian J, et al. Computer-aided detection (CAD) in mammography: does it help the junior or the senior radiologist?[J]. European journal of radiology, 2005, 54(1): 90-96.

[16] Gómez S S, Tabanera M T, Bolivar A V, et al. Impact of a CAD system in a screen-film mammography screening program: A prospective study[J]. European journal of radiology, 2011, 80(3): e317-e321.

[17] Houssein E H, Emam M M, Ali A A, et al. Deep and machine learning techniques for medical imaging-based breast cancer: A comprehensive review[J]. Expert Systems with Applications, 2021, 167: 114161.

[18] D'Orsi C, Bassett L, Feig S. Breast imaging reporting and data system (BI-RADS)[J]. Breast imaging atlas, 4th edn. American College of Radiology, Reston, 2018.

[19] Tot T, Gere M, Hofmeyer S, et al. The clinical value of detecting microcalcifications on a mammogram[C]//Seminars in Cancer Biology. Academic Press, 2021, 72: 165-174.

[20] Nunes F L S, Schiabel H, Goes C E. Contrast enhancement in dense breast images to aid clustered microcalcifications detection[J]. Journal of Digital Imaging, 2007, 20(1): 53-66.

[21] Henriksen E L, Carlsen J F, Vejborg I M M, et al. The efficacy of using computer-aided detection (CAD) for detection of breast cancer in mammography screening: a systematic review[J]. Acta Radiologica, 2019, 60(1): 13-18.

[22] Njeh I, Sassi O B, Chtourou K, et al. Speckle noise reduction in breast ultrasound images: SMU (SRAD median unsharp) approch[C]//Eighth International Multi-Conference on Systems, Signals & Devices. IEEE, 2011: 1-6.

[23] Bai Y, Zhang Y, Ding M, et al. Sod-mtgan: Small object detection via multi-task generative adversarial network[C]//Proceedings of the European Conference on Computer Vision (ECCV). 2018: 206-221.

[24] Kim D W, Jang H Y, Kim K W, et al. Design characteristics of studies reporting the performance of artificial intelligence algorithms for diagnostic analysis of medical

images: results from recently published papers[J]. Korean journal of radiology, 2019, 20(3): 405-410.

[25] Davies D H, Dance D R. Automatic computer detection of clustered calcifications in digital mammograms[J]. Physics in Medicine & Biology, 1990, 35(8): 1111.

[26] Oliver A, Torrent A, Lladó X, et al. Automatic microcalcification and cluster detection for digital and digitised mammograms[J]. Knowledge-Based Systems, 2012, 28: 68-75.

[27] Karahaliou A, Skiadopoulos S, Boniatis I, et al. Texture analysis of tissue surrounding microcalcifications on mammograms for breast cancer diagnosis[J]. The British journal of radiology, 2007, 80(956): 648-656.

[28] Yassin N I R, Omran S, El Houby E M F, et al. Machine learning techniques for breast cancer computer aided diagnosis using different image modalities: A systematic review[J]. Computer methods and programs in biomedicine, 2018, 156: 25-45.

[29] Phadke A C, Rege P P. Fusion of local and global features for classification of abnormality in mammograms[J]. Sādhanā, 2016, 41(4): 385-395.

[30] Litjens G, Kooi T, Bejnordi B E, et al. A survey on deep learning in medical image analysis[J]. Medical image analysis, 2017, 42: 60-88.

[31] Zhang F, Luo L, Sun X, et al. Cascaded generative and discriminative learning for microcalcification detection in breast mammograms[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 12578-12586.

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

[33] Akselrod-Ballin A, Karlinsky L, Hazan A, et al. Deep learning for automatic detection of abnormal findings in breast mammography[M]//Deep learning in medical image analysis and multimodal learning for clinical decision support. Springer, Cham, 2017: 321-329.

[34] Cao Z, Yang Z, Zhuo X, et al. Deeplima: Deep learning based lesion identification in mammograms[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops. 2019: 0-0.

[35] Anwar S M, Majid M, Qayyum A, et al. Medical image analysis using convolutional neural networks: a review[J]. Journal of medical systems, 2018, 42(11): 1-13.

[36] Cheplygina V, de Bruijne M, Pluim J P W. Not-so-supervised: a survey of semi-supervised, multi-instance, and transfer learning in medical image analysis[J]. Medical image analysis, 2019, 54: 280-296.

[37] Agarwal R, Díaz O, Yap M H, et al. Deep learning for mass detection in Full Field Digital Mammograms[J]. Computers in biology and medicine, 2020, 121: 103774.

[38] Shen L, Margolies L R, Rothstein J H, et al. Deep learning to improve breast cancer detection on screening mammography[J]. Scientific reports, 2019, 9(1): 1-12.

[39] Shen Y, Wu N, Phang J, et al. Globally-aware multiple instance classifier for breast cancer screening[C]//International Workshop on Machine Learning in Medical Imaging. Springer, Cham, 2019: 18-26.

[40] Shao F, Chen L, Shao J, et al. Deep Learning for Weakly-Supervised Object Detection and Object Localization: A Survey[J]. arXiv preprint arXiv:2105.12694, 2021.

[41] Liang G, Wang X, Zhang Y, et al. Weakly-supervised self-training for breast cancer localization[C]//2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). IEEE, 2020: 1124-1127.

[42] Samulski M, Karssemeijer N. Optimizing case-based detection performance in a multiview CAD system for mammography[J]. IEEE Transactions on Medical Imaging, 2011, 30(4): 1001-1009.

[43] Bekker A J, Shalhon M, Greenspan H, et al. Multi-view probabilistic classification of breast microcalcifications[J]. IEEE Transactions on medical imaging, 2015, 35(2): 645-653.

[44] Ma J, Li X, Li H, et al. Cross-view relation networks for mammogram mass detection[C]//2020 25th International Conference on Pattern Recognition (ICPR). IEEE, 2021: 8632-8638.

[45] Liu Y, Zhang F, Zhang Q, et al. Cross-view correspondence reasoning based on bipartite graph convolutional network for mammogram mass detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 3812-3822.

[46] Arevalo J, González F A, Ramos-Pollán R, et al. Representation learning for mammography mass lesion classification with convolutional neural networks[J]. Computer methods and programs in biomedicine, 2016, 127: 248-257.

[47] Qiu Y, Yan S, Tan M, et al. Computer-aided classification of mammographic masses using the deep learning technology: a preliminary study[C]//Medical Imaging 2016: Computer-Aided Diagnosis. SPIE, 2016, 9785: 511-516.

[48] Sun W, Tseng T L B, Zhang J, et al. Enhancing deep convolutional neural network scheme for breast cancer diagnosis with unlabeled data[J]. Computerized Medical Imaging and Graphics, 2017, 57: 4-9.

[49] Morrell S, Wojna Z, Khoo C S, et al. Large-scale mammography CAD with deformable conv-nets[M]//Image Analysis for Moving Organ, Breast, and Thoracic Images. Springer, Cham, 2018: 64-72.

[50] Carneiro G, Nascimento J, Bradley A P. Unregistered multiview mammogram analysis with pre-trained deep learning models[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham, 2015: 652-660.

[51] Dhungel N, Carneiro G, Bradley A P. Fully automated classification of mammograms using deep residual neural networks[C]//2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017). IEEE, 2017: 310-314.

[52] Zhao X, Yu L, Wang X. Cross-view attention network for breast cancer screening from multi-view mammograms[C]//ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2020: 1050-1054.

[53] Shen L, Margolies L R, Rothstein J H, et al. Deep learning to improve breast cancer detection on screening mammography[J]. Scientific reports, 2019, 9(1): 1-12.

[54] Shams S, Platania R, Zhang J, et al. Deep generative breast cancer screening and diagnosis[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham, 2018: 859-867.

[55] Hu T, Zhang L, Xie L, et al. A multi-instance networks with multiple views for classification of mammograms[J]. Neurocomputing, 2021, 443: 320-328.

[56] McKinney S M, Sieniek M, Godbole V, et al. International evaluation of an AI system for breast cancer screening[J]. Nature, 2020, 577(7788): 89-94.

[57] Moreira I C, Amaral I, Domingues I, et al. Inbreast: toward a full-field digital mammographic database[J]. Academic radiology, 2012, 19(2): 236-248.

[58] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 2117-2125.

[59] Redmon J, Farhadi A. Yolov3: An incremental improvement[J]. arXiv preprint arXiv:1804.02767, 2018.

[60] Lu Z, Carneiro G, Dhungel N, et al. Automated detection of individual micro-calcifications from mammograms using a multi-stage cascade approach[J]. arXiv preprint arXiv:1610.02251, 2016.

[61] Liu X, Mei M, Liu J, et al. Microcalcification detection in full-field digital mammograms with PFCM clustering and weighted SVM-based method[J]. EURASIP Journal on Advances in Signal Processing, 2015, 2015(1): 1-13.

[62] Illingworth J, Kittler J. A survey of the Hough transform[J]. Computer vision, graphics, and image processing, 1988, 44(1): 87-116.

[63] Tan M, Le Q. Efficientnetv2: Smaller models and faster training[C]//International Conference on Machine Learning. PMLR, 2021: 10096-10106.

[64] Muhammad M B, Yeasin M. Eigen-cam: Class activation map using principal components[C]//2020 International Joint Conference on Neural Networks (IJCNN). IEEE, 2020: 1-7.

[65] Lotter W, Diab A R, Haslam B, et al. Robust breast cancer detection in mammography and digital breast tomosynthesis using an annotation-efficient deep learning approach[J]. Nature Medicine, 2021, 27(2): 244-249.

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

[67] Agarwal R, Diaz O, Lladó X, et al. Automatic mass detection in mammograms using deep convolutional neural networks[J]. Journal of Medical Imaging, 2019, 6(3): 031409.

[68] Yang Z, Cao Z, Zhang Y, et al. MommiNet-v2: Mammographic multi-view mass identification networks[J]. Medical Image Analysis, 2021, 73: 102204.

[69] Tang Y, Cao Z, Zhang Y, et al. Leveraging Large-Scale Weakly Labeled Data for Semi-Supervised Mass Detection in Mammograms[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 3855-3864.

[70] Tardy M, Mateus D. Looking for abnormalities in mammograms with self-and weakly supervised reconstruction[J]. IEEE Transactions on Medical Imaging, 2021, 40(10): 2711-2722.

[71] Ultraytics. YOLOV5[CP/OL]. 2020. https://github.com/ultralytics/YOLOv5.

[72] Shen Y, Wu N, Phang J, et al. An interpretable classifier for high-resolution breast cancer screening images utilizing weakly supervised localization[J]. Medical image analysis, 2021, 68: 101908.

[73] Selvaraju R R, Cogswell M, Das A, et al. Grad-cam: Visual explanations from deep networks via gradient-based localization[C]//Proceedings of the IEEE international conference on computer vision. 2017: 618-626.

[74] Wang H, Wang Z, Du M, et al. Score-CAM: Score-weighted visual explanations for convolutional neural networks[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops. 2020: 24-25.