Towards improved breast mass detection using dual-view mammogram matching

Yutong Yan, Pierre-Henri Conze, Mathieu Lamard, Gwenolé Quellec, Béatrice Cochener, Gouenou Coatrieux

Medical Image Analysis (2021)

Contents lists available at ScienceDirect

Medical Image Analysis

journal homepage: www.elsevier.com/locate/media

Towards improved breast mass detection using dual-view mammogram matching

Yutong Yan[a,b,c], Pierre-Henri Conze[a,c,*], Mathieu Lamard[a,b], Gwenolé Quellec[a], Béatrice Cochener[a,c,d], Gouenou Coatrieux[a,c]

[a] *Inserm, LaTIM UMR 1101, 22 rue Camille Desmoulins, 29238 Brest, France*
[b] *Université de Bretagne Occidentale, 3 rue des Archives, 29238 Brest, France*
[c] *IMT Atlantique, Technopôle Brest-Iroise, 29238 Brest, France*
[d] *CHRU de Brest, 2 avenue Foch, 29200 Brest, France*

ARTICLE INFO

Article history:
Received March 2020
Received in final form -
Accepted -
Available online -

Communicated by -

Keywords: breast cancer, computer-aided diagnosis, mass detection, dual-view matching, siamese networks, information fusion

ABSTRACT

Breast cancer screening benefits from the visual analysis of multiple views of routine mammograms. As for clinical practice, computer-aided diagnosis (CAD) systems could be enhanced by integrating multi-view information. In this work, we propose a new multi-tasking framework that combines craniocaudal (CC) and mediolateral-oblique (MLO) mammograms for automatic breast mass detection. Rather than addressing mass recognition only, we exploit multi-tasking properties of deep networks to jointly learn mass matching and classification, towards better detection performance. Specifically, we propose a unified Siamese network that combines patch-level mass/non-mass classification and dual-view mass matching to take full advantage of multi-view information. This model is exploited in a full image detection pipeline based on You-Only-Look-Once (YOLO) region proposals. We carry out exhaustive experiments to highlight the contribution of dual-view matching for both patch-level classification and examination-level detection scenarios. Results demonstrate that mass matching highly improves the full-pipeline detection performance by outperforming conventional single-task schemes with 94.78% as Area Under the Curve (AUC) score and a classification accuracy of 0.8791. Interestingly, mass classification also improves the performance of mass matching, which proves the complementarity of both tasks. Our method further guides clinicians by providing accurate dual-view mass correspondences, which suggests that it could act as a relevant second opinion for mammogram interpretation and breast cancer diagnosis.

1. Introduction

Breast cancer is the second most common cause of cancer-related deaths in women (Rakhlin et al., 2018), responsible for 25% of cancer cases and 15% of cancer deaths (Torre et al., 2017). Mammography is the main imaging modality used to detect breast abnormalities at early stage. Standard mammography views are bilateral craniocaudal (CC), extracted from top-down, and mediolateral-oblique (MLO), an oblique view taken under 45°. These two views comprise routine screening mammography. Dual-view analysis is an effective way for clinicians to reduce both morbidity and mortality associated with breast cancer (Jørgensen and Bewley, 2015). Examining the correspondence between suspicious findings in multiples views enables radiologists to improve interpretations (Vijayarajan and Jaganathan, 2014). Compared to single-view screening, the clinical use of dual-view techniques therefore reduces false-positive cases and improves cancer detection rates (Warren et al., 1996). This analysis is mainly related to the de-

*Corresponding author: Tel.: +33-229-001-411;
E-mail: pierre-henri.conze@imt-atlantique.fr (P.-H. Conze)

tection and classification of lesions such as masses, calcifications, asymmetry or distortions. Among those abnormalities, masses are the most important clinical symptoms of carcinomas. Characterized by medium gray to white regions within the breast area, masses exhibit a great diversity of size, shape (irregular, oval, lobulated, round), contours (circumscribed, ill-defined, spiculated, obscured) and texture (Yan et al., 2019).

Mammogram analysis is usually performed manually by a radiologist. This task is time-consuming and prone to strong inter-expert variability (Hamidinekoo et al., 2018). Moreover, it is difficult and impractical for clinicians to perform double reading in most screening situations. This leads to a considerable amount of patients which are given heavy treatments by mistake (Myers et al., 2015). Computer-aided diagnosis (CAD) systems have been designed for supplemental lesion detection, classification and segmentation purposes. However, conventional CAD systems for mammogram interpretation are inefficient and not automatic enough to significantly improve diagnosis performance (Lehman et al., 2015). The use of multi-view contexts is a known weakness of current CAD technology.

In recent years, deep learning has achieved remarkable breakthroughs in medical image analysis through convolutional neural networks (CNN). Deep models have shown the most promising performance in recent breast cancer mammography-related competitions (Hamidinekoo et al., 2018). CAD systems that employ deep learning demonstrate stronger robustness in clinical implementation than traditional methodologies. The main advantage lies in avoiding the need of hand-crafted features and automatically learning representative features directly from data. Nevertheless, breast mass detection and classification are still open issues due to the strong variations in mass appearance (Yan et al., 2019). It remains therefore difficult to distinguish a mass from its surrounding healthy tissues. Some studies (Geras et al., 2017; Zhu et al., 2017; Zhang et al., 2018; Shen et al., 2019) focus on whole mammograms which simplify such complex problem by providing a unique image-level label (normal, benign or malignant). The drawback is that it avoids conducting a comprehensive analysis comprising lesion types and locations. Other works are mostly region-based methods (Wang et al., 2018; Choukroun et al., 2017; Arevalo et al., 2015; Lévy and Jain, 2016; Zhou et al., 2017), where images are first decomposed into regions to further distinguish normal from abnormal tissues. However, most of the above methods use single-view mammograms only, thus neglecting the rich information that can be extracted from multi-view images.

To address the limitation of single-view processing, we aim at taking advantage of information arising from CC and MLO mammograms, as do clinicians when making decisions in clinical practice (Vijayarajan and Jaganathan, 2014). There is a huge potential to improve the performance of CAD systems by integrating information from paired views. The concept of multi-view information fusion was recently introduced to improve the performance of detection, classification or content-based mammogram retrieval tasks (Jouirou et al., 2019). Several multi-view fusion schemes learn on full images from each view separately and concatenate respective features afterwards. (Geras et al., 2017) proposed to apply CNN models to each view separately to obtain view-specific representations for further classification purposes. Nevertheless, such late-fusion schemes only exploit image-level view-specific representations.

Alternatively, we propose a novel multi-tasking Siamese deep model that combines CC and MLO mammograms to improve breast mass detection. A Siamese model includes two identical sub-networks with shared weights, such that features from two different input images can be extracted simultaneously. Previous related works (Ma et al., 2019; Perek et al., 2018) also employ Siamese networks (Koch et al., 2015) for multi-view study. However, these are single-task studies dedicated to mass detection (Ma et al., 2019) or mass matching (Perek et al., 2018) only. To design a more comprehensive and efficient CAD system, we aim at exploiting the multi-tasking properties of deep CNN. Multi-task learning processes multiple tasks jointly with many advantages such as saving computation time and resources as well as improving robustness against overfitting (Ruder, 2017). The network parameters from feature extraction layers are updated through the optimization of a combined loss dealing with both mass/non-mass classification and matching. Contrary to (Ma et al., 2019; Perek et al., 2018), our method can provide both classification and matching results. Specifically, our contributions are two-folds. First, we propose a new deep learning algorithm that capitalizes on multi-view fusion and multi-task learning to improve breast mass detection. To the best of our knowledge, our framework is the first that exploits multi-tasking abilities of deep learning models to improve mass detection using multi-view matching. Second, we conduct a comprehensive evaluation of various networks towards multi-task learning on public datasets. Both quantitative and visual results prove the effectiveness of the proposed strategy.

This paper is organized as follows. We present background material and previous works related to mass classification, detection and matching using deep learning in Sect.2. Sect.3 focuses on our methodology for jointly learning mass/non-mass classification and matching. Sect.4 reports and discusses experimental results extending a preliminary study presented in (Yan et al., 2020a). We end up with conclusions in Sect.5.

2. Related works

2.1. Mass detection and classification

Deep learning methods have shown impressive performance in medical image applications. CNN models make it possible to automatically extract and learn features without feature engineering. Several works have started to explore deep learning for mass detection and classification. Arevalo et al. (2015) applied a hybrid approach in which CNNs are used to learn the feature representation in a supervised way. Zhou et al. (2017) analyzed the effectiveness of CNN in abnormality detection and classified lesions as benign or malignant. Lévy and Jain (2016) evaluated three different CNN architectures including shallow CNN, AlexNet (Krizhevsky et al., 2012) and GoogLeNet (Szegedy et al., 2015) for breast mass classification and further studied the performance reached by each architecture. Afterwards, Zhu et al. (2017) conducted benign/malign

mass classification based on whole mammograms using an end-to-end deep multi-instance network without any detection or segmentation annotations for training. Dhungel et al. (2017) proposed an integrated cascade of deep belief networks and gaussian mixture models to provide mass candidates, followed by cascades of CNN and random forest classifiers to refine detection results. Agarwal et al. (2019) developed an end-to-end mass detection framework using a patch-level classification approach. However, the above methods only exploit single-view mammograms, and therefore neglect the rich information arising from multi-view images which appears as essential to improve breast cancer diagnosis.

2.2. Multi-view mammography analysis

To address the single-view interpretation issue, a growing number of works focus on multi-view information fusion. Vijayarajan and Jaganathan (2014) extracted 2D features from whole mammograms, obtained the locations from CC and MLO views and merged this information to get a 3D view of the mass location. Carneiro et al. (2015) trained a separate CNN model for each view and finally applied a CNN classifier that estimates the BI-RADS score using features learned from unregistered CC and MLO mammograms, as well as respective mass delineations. Geras et al. (2017) proposed to apply a CNN model separately to each view to obtain view-specific representations for further classification purposes. All the above studies are designed based on whole mammograms. However, there may be multiple different benign or malignant masses in a given examination. In order to simplify the complex analysis of whole mammograms, some studies assign a unique label (benign, malignant or normal) to the whole image. The drawback is that it avoids conducting a comprehensive analysis of each mammogram, comprising lesion types and locations.

2.3. Patch-based matching

Patch-based image matching has been extensively used in computer vision. Han et al. (2015) presented MatchNet, a deep convolutional approach based on Siamese networks for patch-based matching between two images I_1 and I_2. The MatchNet architecture consists of a feature network followed by a metric network. The former is a "two-tower" structure network which jointly processes two patches (one extracted from I_1, another from I_2) and maps them to a feature representation. The latter models the similarity between the paired features through fully-connected (FC) layers and a softmax layer to get a matching score. Zagoruyko and Komodakis (2015) made use of and explored CNN architectures to encode a general similarity function to quantify the correspondence between image patches. Amit et al. (2015) combined unsupervised segmentation and random-forest classification to detect candidate masses in CC and MLO views before estimating the correspondence between pairs of candidates in the two views. Ma et al. (2019) exploited the latent relation information between the corresponding mass regions of interest (RoI) from the two paired views using a cross-view relation region-based CNN for mass detection. Based on MatchNet (Han et al., 2015), Perek et al. (2018) proposed a dual-view Siamese network that learns patch representations and similarity for lesion matching. This suggests a potential added value of multi-view matching to improve breast mass detection, with respect to single-view detection strategies.

3. Methods

We first formally define the problem settings and provide an overview of the proposed unified framework for mass classification and matching in Sect.3.1. Multi-view mass matching combining Siamese networks and contrastive learning is described in Sect.3.2. As presented in Sect.3.3, this methodology is then extended to address both patch matching and classification simultaneously. Multi-task learning is followed to obtain better predictive breast mass classification performance than traditional single-task learning schemes.

3.1. Overview

Our multi-tasking framework (Fig.1) takes unregistered CC/MLO view pairs as inputs and provides as outputs accurate mass detections along with correspondences between mass regions in both views. Among existing deep detectors including Faster R-CNN (Ren et al., 2015) or SSD (Liu et al., 2016), YOLOv3 (Redmon and Farhadi, 2018) is adopted for candidate patch generation and selection from high resolution full mammograms (Yan et al., 2020b), since it offers an excellent trade-off between accuracy and efficiency.

Given a pair of mammograms $\{I_{CC}, I_{MLO}\}$, YOLO predicts two sets of candidate mass patches $P_{CC} = \{p^1_{CC}, ..., p^N_{CC}\}$ and $P_{MLO} = \{p^1_{MLO}, ..., p^M_{MLO}\}$. Although recent deep learning-based detectors have yielded impressive accuracy for object detection in natural images, it still remains difficult to reach the same level of performance when applied to medical images, especially mammograms. The following reasons arise. First, object size variance may affect the performance. In our context, mass sizes and aspect ratios vary strongly (Yan et al., 2019). Second, mass detection is generally more difficult than common object detection since masses are visually less obvious and less contrasted with respect to their surrounding healthy tissues, combined with a great diversity of shape and texture. On top of that, we should also struggle with the barrier between true and false masses to retain as much as possible true positives while reducing false positives.

To further finely select mass candidates and discover the latent relation between CC and MLO views, we design a combined model through a Siamese network that jointly deals with patch-level mass/non-mass classification and matching (Fig.1). We sample candidate mass patches P_{CC} and P_{MLO} to the same size via a data sampler, while performing data augmentation to prevent from overfitting. These samples are then fed into our combined network. Based on robust generic feature extraction, the result of our model is whether each patch of the two views contains mass as well as the correspondence between two patches of two views. Subsequently, we can visualize final detection results on both views to further guide clinicians in their mammogram interpretation task.

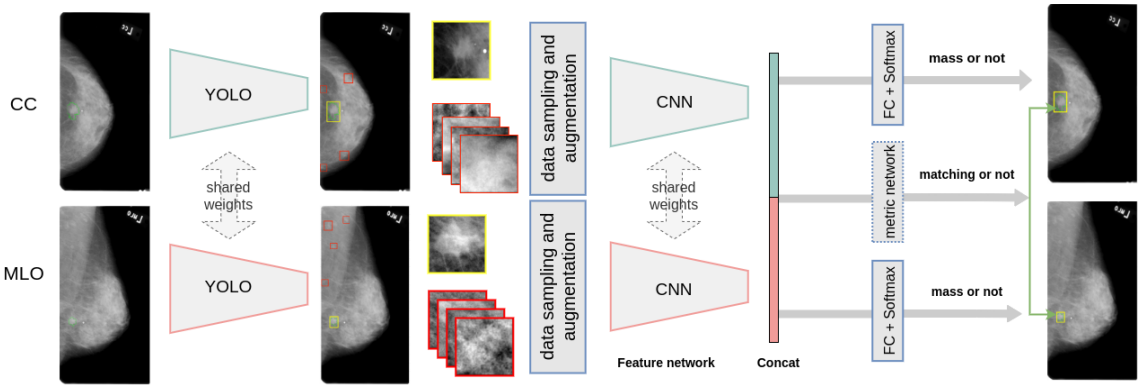

Fig. 1. Proposed multi-tasking deep pipeline. In images, green contours indicate ground truth delineations, red (yellow) boxes indicate false (true) detections.

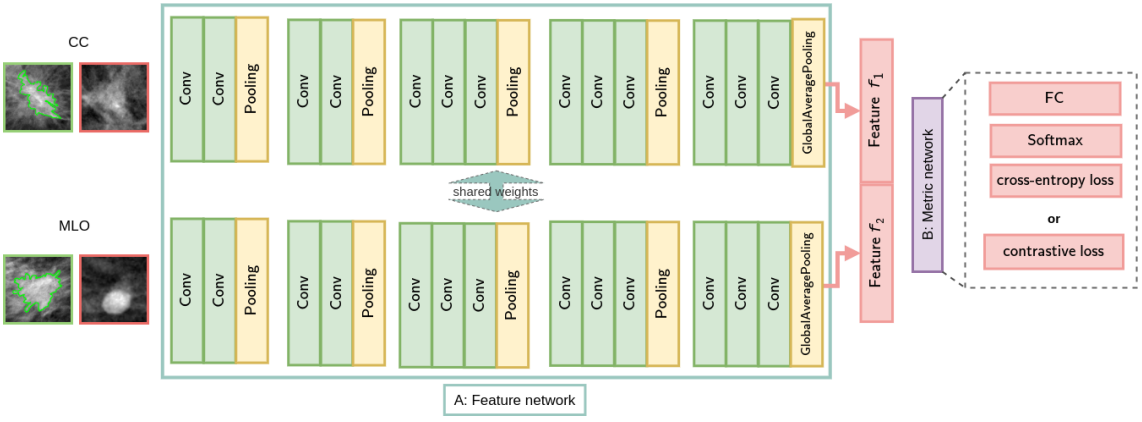

Fig. 2. Matching Siamese network. A: Two-branch feature network which takes as input both positive (green patch) and negative (red patch) patch samples of CC and MLO views separately to compute features. Resulting features f_1 and f_2 are concatenated for patch comparison. B: Metric network.

3.2. *Dual-view mammogram matching*

Essential to make decisions in clinical routine, the information presented in the two paired CC/MLO views is highly complementary and could serve as a second source of decision (Jouirou et al., 2019). However, due to breast deformation and different acquisition conditions combined with the lack of 3D information, multi-view fusion for dual-view mammogram analysis is a challenging task. Therefore, only few deep convolutional methods for breast screening consider learning jointly effective features from both views.

Inspired by (Perek et al., 2018) and (Han et al., 2015), we employ a Siamese framework to identify correspondences between masses in both CC/MLO views. The deep architecture for multi-view mammogram matching is shown in Fig.2. Patch pairs from CC and MLO views are fed separately to the two branches of the network. The feature network A is a Siamese model in which two fully convolutional networks with shared weights are employed for feature extraction. For illustration (Fig.2), we use a VGG16 architecture (Simonyan and Zisserman, 2014) with repeated 3×3 convolutions followed by an activation function (ReLU) and 2×2 max pooling. To reduce the number of parameters and avoid overfitting, we apply a global average pooling layer before subsequent FC layers (Szegedy et al., 2015; He et al., 2016). Particularly, different widely used deep convolutional models such as VGG16 (Simonyan and Zisserman, 2014), ResNet50, ResNet101 (He et al., 2016), InceptionV3 (Szegedy et al., 2016) and EfficientNet (Tan and Le, 2019) can be exploited for feature extraction purposes. For feature comparison, two manners are explored based on different loss functions. First, one can use a metric network as in (Perek et al., 2018) and (Han et al., 2015) consisting of several FC layers and softmax layers, trained with a cross-entropy loss. Alternatively, we can rather employ a contrastive loss (Hadsell et al., 2006) to improve the representation ability of network A to extract discriminative features.

Contrastive loss for matching. Contrastive learning, whose labels are used to guide the choice of positive and negative pairs, is employed to learn powerful feature representations. The contrastive loss is usually exploited for image retrieval tasks, along with Siamese networks to learn paired data relationships. During training, an image pair is fed into the model with their ground truth relationship Y. The loss function is as follows:

$$L_{mat}(Y, X_1, X_2) = \frac{1}{2N} \sum_{n=1}^{N} Y D_W^2 + (1 - Y) \max(m - D_W, 0)^2 \quad (1)$$

where $D_W(X_1, X_2) = \|f_1 - f_2\|_2$ represents the Euclidean distance between two sample features f_1 and f_2. Y is the label of whether the two samples match. $Y = 1$ if the two samples

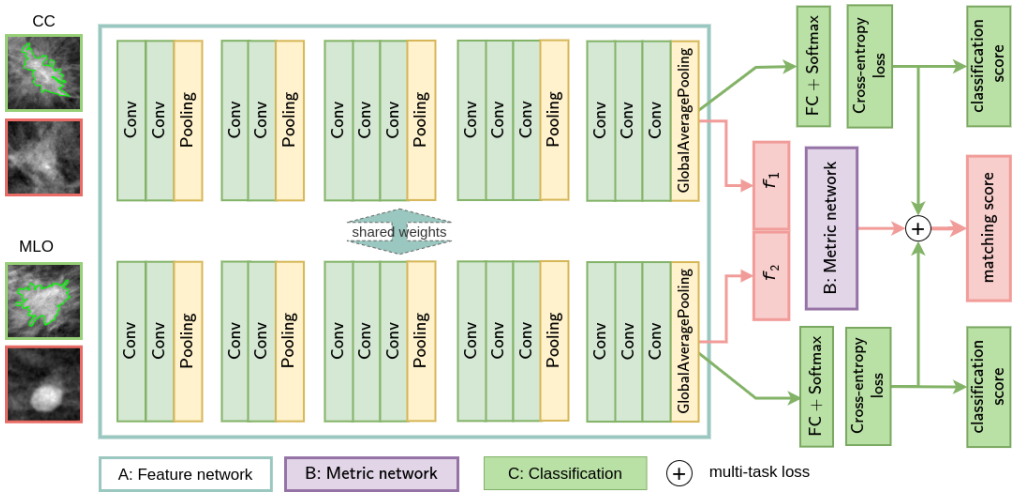

Fig. 3. The proposed Combined Matching and Classification Network (CMCNet). Green (red) patches correspond to positive (negative) samples.

are similar and 0 otherwise. $m > 0$ is a margin that defines a radius: dissimilar pairs contribute to the loss only if their distance is within this radius. N is the number of samples. Unlike conventional learning systems where the loss function is a sum over samples, the contrastive loss runs over pairs of feature vectors $\{f_1, f_2\}$ such that there is no more need for FC and softmax layers. Moreover, compared to cross-entropy which learns the patch "match" or "not match" in an inexplicable manner, the contrastive loss optimizes the mass matching task by manipulating the distance between pairs in feature space. Therefore, the contrastive loss is more in line with matching requirements than binary sample classification. The loss function L_{mat} (Eq.1) is minimized using stochastic gradient descent (SGD).

3.3. Combined mass matching and classification

Mass classification and dual-view matching are two tasks of a very different nature. The challenge is thus to learn generic features for both tasks. We propose to exploit Siamese networks towards simultaneous deep patch-level matching and classification. In this direction, we design a multi-tasking learning model (Fig.3) referred as Combined Matching and Classification Network (CMCNet). Positive and negative patch samples of CC/MLO views arising from YOLOv3 detector are fed into the two-branch feature network (Fig.3-A) to compute robust patch representations. Apart from the matching network (Sect.3.2, Fig.3-B), we incorporate into the pipeline two branches (Fig.3-C) for CC/MLO mass classification purposes. Each of these branches has its own FC layers. We not only jointly learn representations from the two views but also simultaneously learn matching and classification tasks to exploit the potential relationship between view-points.

The combined learning of classification and matching refers to the idea of multi-task learning which has been proven to improve learning efficiency and generalization performance of task-specific models. We expect thus that the dual-view matching task can improve the robustness of mass classification, towards better predictive results than classification-only strategies. The designed loss L is the sum of three losses to optimize the entire CMCNet parameters through SGD:

$$L = \alpha L_{cls,CC} + \beta L_{cls,MLO} + \gamma L_{mat} \quad (2)$$

where $L_{cls,CC}$ and $L_{cls,MLO}$ represent the classification loss (cross-entropy) for CC and MLO view respectively. L_{mat} is the matching loss which can be cross-entropy or contrastive loss (Eq.1). α, β and γ are coefficients balancing the loss terms.

4. Experiments and results

In what follows, we present the datasets used in this study in Sect.4.1. Sect.4.2 reports the experimental settings. We evaluate the proposed approach both quantitatively and qualitatively in Sect.4.3. Deep models for patch classification and matching are implemented using pytorch. Experiments are performed with a single Nvidia GeForce GTX 1080Ti GPU (11GB/s).

4.1. Imaging datasets

We focus on mass detection from high-resolution 2048×1024 mammograms arising from two publicly-available mammogram datasets: INbreast (Moreira et al., 2012) and DDSM-CBIS (Lee et al., 2017). Here are the dataset characteristics:

- INbreast (Moreira et al., 2012) has a total of 410 mammograms from 115 examinations, with four images per case corresponding to the four standard views used in screening mammography: R-CC, L-CC, R-MLO, and L-MLO where L and R stand respectively for left and right. Four types of lesions (masses, calcifications, asymmetry and distortions) are included but only 107 images contain masses for which accurate contours made by specialists are provided. In this work, we selected 35 CC/MLO pairs (70 images) out of 107 examinations with mass.

- DDSM-CBIS (Digital Database for Screening Mammography) (Lee et al., 2017) is a larger database which contains approximately 2,500 mammograms including normal, benign, and malignant cases with verified pathology information and coarse ground truth manual delineations.

	DDSM-CBIS		INbreast	
	training	validation	test	full-pipeline
classif.	4690/4690	1170/1170	700/700	125/225
matching	2345/4690	585/1170	350/700	125/225

Table 1. Data distribution setting for experiments. Each cell has the following format: number of positive samplings / number of negative samplings.

Network	learning rate	batch size	margin (m)	γ
VGG16	0.0005	128	15	0.1
ResNet50	0.0005	128	15	0.1
ResNet101	0.001	64	10	0.1
InceptionV3	0.001	32	10	0.1
EfficientNet-B3	0.001	128	10	0.1

Table 2. Optimal hyper-parameters used for each deep network.

In this work, we selected 586 pairs of CC/MLO mammograms (1172 images).

During training, the principle followed for CC/MLO pair selection is that both views should contain only one mass. The reason is that if the image contains two or more masses, multi-view correspondences become unknown without the help of an expert. Accordingly, the existence of a single pair is the only data selection criterion for training. However, this is not a limitation for inference or clinical use. Among the 586 DDSM-CBIS pairs, 80% are used for training and the remaining 20% for validation. The 35 INbreast pairs are only used in the testing stage since it is too small to be representative as training data. The only pre-processing step is to crop most of the blank area from the original image (4084×3328 or 3328×2560) and resize the remaining area to 2048×1024.

4.2. Experimental settings

Data sampling and augmentation. In our multi-task framework, sampling in training is crucial. However, regions of healthy tissues in a whole mammogram are much larger than the mass areas, leading to inevitable false positive YOLOv3 proposals. Similarly, sample imbalance can make the deep classification model very biased. To minimize these effects, data sampling is conducted as follows. For classification training, positive samples are taken according to provided ground truth masks, while negative patches are generated by YOLOv3 (Sect.3.1). In particular, we randomly generate K patches per image with an intersection over union (IoU) with respect to the ground truth box larger than 0.5. In practice, $K = 5$ (respectively 10) for DDSM-CBIS (INbreast) since the INbreast dataset is much smaller. Likewise, we choose K negative patches from false YOLO predictions. We thus use a very small threshold ($<10^{-4}$) on detection probabilities to retain as many predictions as possible and select the K false candidates with the highest scores. All patches are resized to 64×64 pixels, as in Han et al. (2015) and Perek et al. (2018). Random rotations of 25 degree, random horizontal flips and random resized crops are applied for data augmentation. For matching, we consider a pair of positive patches of the same mass from the two views as a matching sample. If one of the patches is labeled negative, they are considered as a negative match. The detailed data distribution is shown in Tab.1 for both DDSM-CBIS and INbreast datasets.

Training patch-level classification and matching. As a proof of concept, we conduct experiments using various model backbones for the feature network: VGG16 (Simonyan and Zisserman, 2014), ResNet50, ResNet101 (He et al., 2016), InceptionV3 (Szegedy et al., 2016) and EfficientNet (Tan and Le, 2019). The feature size varies depending on the model used. Let M be the number of feature map channels and B denote the batch size. The FC layers of each classification branch will turn the input vector (B, M) into $(B, 2)$ and pass it to the Softmax layer to transfer logits into probabilistic predictions. The input of the metric network (Fig.3-B) is the concatenation of two feature vectors. For VGG16, $M = 512$. For ResNet50 and ResNet101, $M = 2048$. For EfficientNet, it has 8 pre-trained models from EfficientNet-B0 to B7 where M is respectively $\{1280, 1280, 1408, 1536, 1792, 2048, 2304, 2560\}$. All deep models are initialized using pre-trained weights (Litjens et al., 2017) from the ImageNet dataset (Russakovsky et al., 2015) and trained using the SGD optimizer. Optimal hyper-parameters vary depending on the network. For loss functions, we choose $\alpha = \beta = 1$ and $\gamma = 1$ for cross-entropy and $\gamma = 0.1$ and margin $m \in \{5, 10, 15\}$ for contrastive loss (Eq.1). The detailed hyper parameters used are shown in Tab.2.

4.3. Evaluation on clinical data

Multi-task learning versus classification-only. Classification performances are measured using classification accuracy (`acc`). We calculate the accuracy of each view separately and collectively. The statistical significance of the multi-tasking model with respect to the classification-only baseline is estimated using Student's t-tests (Tab.3). Overall, we observe better classification results on MLO than on CC views. In most cases, multi-tasking models that combine classification and matching are better than classification-only from 2% to 4% in accuracy with statistical significance ($p < 0.05$), which reflects the benefits of dual-view matching. Except for ResNet101, we obtain slight gains with the contrastive loss compared to cross-entropy. The difference between networks is not obvious. ResNet101 achieves the best overall accuracy with statistical significance (`acc` = 0.9098, p = 0.007 compared to baseline). Improvements obtained by VGG16 using the contrastive loss are also significant (`acc` = 0.9084, $p < 1e^{-6}$), followed by ResNet50 (`acc` = 0.9049), InceptionV3 (`acc` = 0.90) and EfficientNet-B3 (`acc` = 0.8979), showing that using deeper networks is not necessary to reach better performance.

Full detection pipeline. To further prove the effectiveness of our method, we conduct experiments with a full detection pipeline. Instead of extracting positive candidates using ground truth mass delineations while using YOLO as a negative patch generator, we use YOLO to generate all candidate patches.

Specifically, coarse mass YOLO detections (Yan et al., 2020b) are performed on INbreast images to generate testing samples. YOLO is pre-trained on ImageNet and fine-tuned on

Network	matching	matching loss	CC acc	MLO acc	overall acc	p-value
VGG16 (Simonyan and Zisserman, 2014)	×	-	0.8558	0.8857	0.8699	-
	√	cross-entropy	0.8796	**0.9163**	0.8958	$<1e^{-6}$
	√	contrastive	**0.9061**	0.9156	**0.9084**	$<1e^{-6}$
ResNet50 (He et al., 2016)	×	-	0.8517	0.9034	0.8734	-
	√	cross-entropy	0.8958	0.9116	0.9014	$4e^{-4}$
	√	contrastive	**0.9010**	**0.9122**	**0.9049**	$<1e^{-6}$
ResNet101 (He et al., 2016)	×	-	0.8680	0.9097	0.8823	-
	√	cross-entropy	**0.8980**	**0.9265**	**0.9098**	0.007
	√	contrastive	0.8891	0.9252	0.9049	0.003
InceptionV3 (Szegedy et al., 2016)	×	-	0.8238	0.8980	0.8601	-
	√	cross-entropy	0.8776	**0.9184**	0.8972	$<1e^{-6}$
	√	contrastive	**0.8946**	0.9095	**0.9000**	$4e^{-4}$
EfficientNet-B3 (Tan and Le, 2019)	×	-	0.8701	0.8803	0.8741	-
	√	cross-entropy	0.8748	0.9116	0.8923	0.065
	√	contrastive	**0.8830**	**0.9163**	**0.8979**	$<1e^{-6}$

Table 3. Multi-task learning (with cross-entropy and contrastive losses) versus classification-only. Results include CC, MLO and overall classification accuracy (acc) as well as statistical significance p-values with respect to the classification-only baseline. Best results per network are in bold.

1514 DDSM-CBIS images. Thereafter, we use a small threshold (10^{-4}) on detection probabilities to ensure that predictions with high and low confidence are both selected. The averaged inference time per image is 78.7ms. We finally obtain 350 candidates, labeled as positive (125 cases) or negative (225 cases) according to IoU ($\geq$ or $<$ than 0.5) between RoIs and ground truth. All 350 candidate patches arising from INbreast are for the test set and all combinations are evaluated. The performance of each setting (classification-only, CMCNet with cross-entropy and CMCNet with contrastive loss using different deep models) is measured using the AUC (Area Under the receiver operating characteristics Curve).

Results for full-pipeline experiments (Tab.4) show that the classification performance is highly improved over the baseline models by combining dual-view matching. In terms of AUC, the performance of VGG16 (resp. Resnet101) increases from 90.47% (71.46%) to 94.78% (92.82%), which corresponds to a gain of 3.31% (21.36%). These results prove the appropriateness of our contributions. The best AUC score (94.78%, p = 0.001) is obtained using the VGG16 model trained with contrastive loss, with an overall accuracy of 0.8791. Results using the contrastive loss are slightly better than cross-entropy in most cases, except for ResNet50. InceptionV3 using cross-entropy and EfficientNet using both losses improve moderately without statistical significance ($p > 0.05$). Compared to Tab.3, the advantages of combining classification and matching are more highlighted with full-pipeline experiments. Higher AUC indicates that we can significantly reduce false positive proposals resulting from YOLO. We also compute the inference time per image to compare computing time costs of each method (Tab.4). This includes testing all possible pairs. The inference time of the CMCNet varies from 2.7 (VGG16) to 25.4ms (EfficientNet-B3). Since no significant improvement arises when using deeper models, models with low time complexity (VGG, ResNet) are more appropriate. Multi-tasking methods do not cost more time than classification-only schemes. Computing time increases significantly with model complexity, whereas no significant improvement arises. The time increase with respect to YOLO detector (78.7ms per image) is almost negligible. This demonstrates that using very deep networks is not useful for our application.

Additionally, we provide mass matching performances during inference. As shown in Tab.5, mass matching performance is measured using accuracy (*acc*) and AUC. We compare mass matching using our multi-task learning (with cross-entropy and contrastive losses) versus matching-only. The matching-only scheme refers to the matching Siamese network illustrated in Fig.2. Results show that the proposed multi-task learning brings gain from 0.62% to 8.64% in AUC and from 0.9% to 11.12% in *acc*. Best results are achieved by ResNet50 (AUC = 94.30%, *acc* = 89.49). On the basis of the experimental results, we can draw the conclusion that not only matching can improve classification, classification can also improve matching, proving that the multi-tasking properties and the multi-view learning can help towards better breast cancer diagnosis and management.

Using the INbreast dataset, we also compare the overall mass detection performance using the True positive rate (TPR) at the average false positive per image (FPavg) with state-of-the-art methods (Tab.6). Since there is no official split of INbreast, each study has its own split between training, testing and validation subsets. Results shown in the top part of Tab.6 give an idea of the overall detection performance without giving a relevant comparison with these studies. For a fair comparison with respect to state-of-the-art, we re-implemented the recently published method of Agarwal et al. (2019) and conducted experiments using the same data as used in our work (80% DDSM-CBIS for training, 20% DDSM-CBIS for validation and 70 INbreast images for testing) to obtain the Free Response Operating Characteristic (FROC) curve of final detections. The bottom part of Tab.6 includes results obtained on the same testing data. In particular, it displays the best TPR@FPavg score achieved using Agarwal et al. (2019): 0.74@0.99. The best TPR@FPavg score (0.96@0.23) is reached by the proposed framework (CMCNet with VGG16 and contrastive loss). It outperforms the classification-only model (0.89@0.29) and shows consistent performance with respect to existing approaches such as Yan

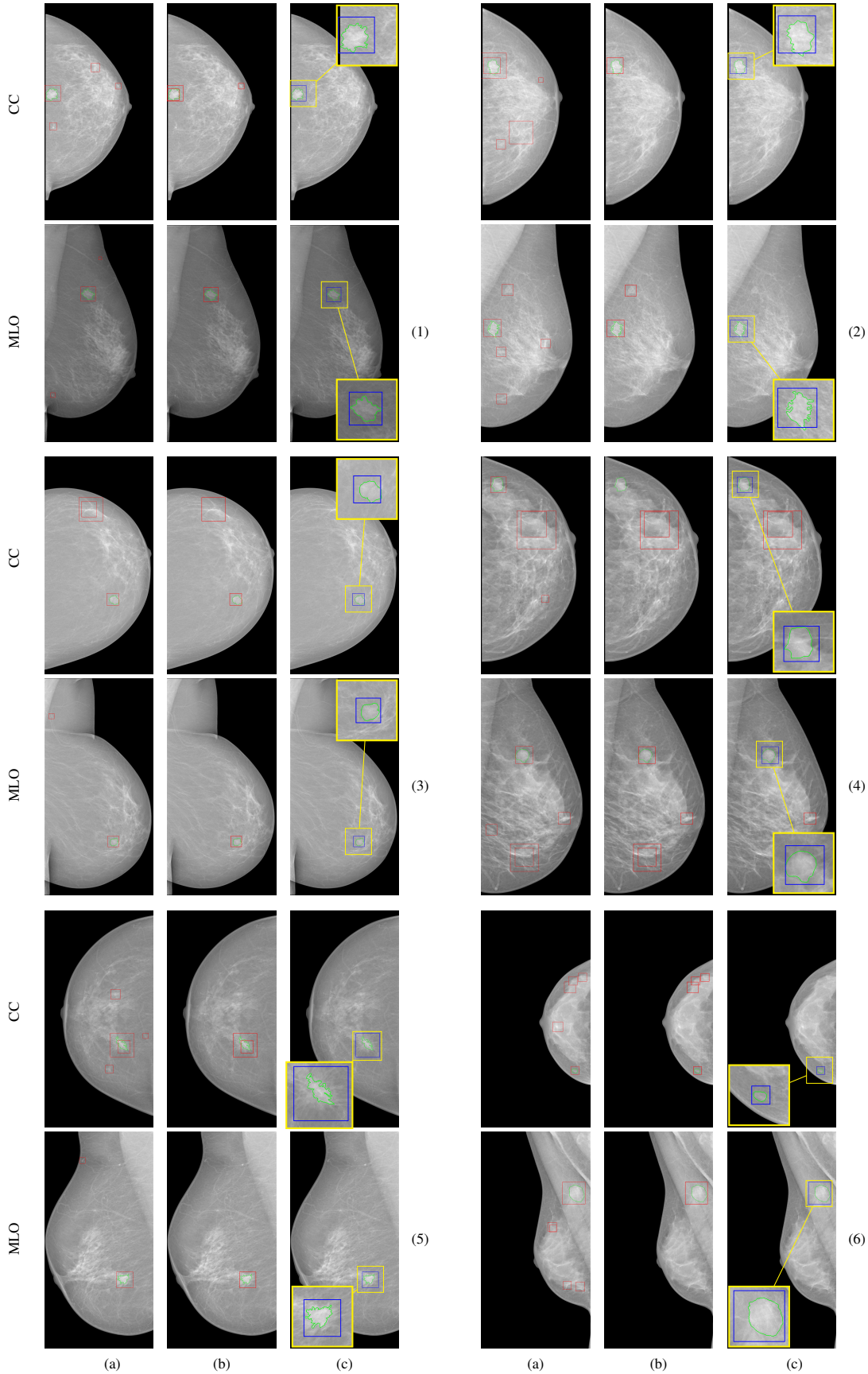

Fig. 4. Full-pipeline mass detection: (a) YOLO detection only, (b) YOLO followed by a classification-only model, (c) YOLO followed by the proposed combined model (with VGG16 and contrastive loss). Red and blue boxes are detected mass bounding boxes. Green labels represent ground truth annotations. Blue boxes show the matching pair selected through dual-view matching. Visual examples are labeled from (1) to (6).

Network	matching	matching loss	overall acc	AUC (%)	AUC p-value	inference time (ms)
VGG16 (Simonyan and Zisserman, 2014)	×	-	0.8260	90.47	-	2.7
	√	cross-entropy	0.8761	94.17	$2e^{-5}$	2.7
	√	contrastive	**0.8791**	**94.78**	0.001	2.7
Resnet50 (He et al., 2016)	×	-	0.6814	70.03	-	8.4
	√	cross-entropy	**0.8555**	**91.98**	$<1e^{-6}$	8.3
	√	contrastive	0.8496	90.30	$<1e^{-6}$	8.4
Resnet101 (He et al., 2016)	×	-	0.7080	71.46	-	16.2
	√	cross-entropy	0.8555	91.74	$<1e^{-6}$	15.8
	√	contrastive	**0.8584**	**92.82**	$<1e^{-6}$	16.3
InceptionV3 (Szegedy et al., 2016)	×	-	0.8112	89.75	-	17.2
	√	cross-entropy	0.8201	89.86	0.9142	16.7
	√	contrastive	**0.8702**	**93.61**	0.009	16.8
EfficientNet-B3 (Tan and Le, 2019)	×	-	0.8142	87.97	-	25.2
	√	cross-entropy	0.8378	**89.80**	0.1795	25.4
	√	contrastive	**0.8466**	88.91	0.5735	24.7

Table 4. Full detection pipeline results including overall classification accuracy (acc), AUC scores, statistical significance p-values of AUC scores with respect to the classification-only baseline, as well as inference times per image. Best results per network are in bold.

Network	matching	classification	matching loss	matching AUC(%)	matching acc
Perek et al. (Perek et al., 2018)	√	×	cross-entropy	79.92	0.7504
VGG16 (Simonyan and Zisserman, 2014)	√	×	cross-entropy	91.05	0.8523
	√	√	cross-entropy	91.49	0.8671
	√	√	contrastive	**92.97**	**0.8714**
ResNet50 (He et al., 2016)	√	×	cross-entropy	92.77	0.8693
	√	√	cross-entropy	92.46	0.8775
	√	√	contrastive	**94.30**	**0.8949**
ResNet101 (He et al., 2016)	√	×	cross-entropy	90.01	0.8345
	√	√	cross-entropy	**92.31**	0.8716
	√	√	contrastive	91.72	**0.8758**
InceptionV3 (Szegedy et al., 2016)	√	×	cross-entropy	89.50	0.8405
	√	√	cross-entropy	**90.64**	0.8536
	√	√	contrastive	90.01	**0.8588**
EfficientNet-B3 (Tan and Le, 2019)	√	×	cross-entropy	82.99	0.7391
	√	√	cross-entropy	89.50	0.8379
	√	√	contrastive	**91.63**	**0.8503**

Table 5. Mass matching AUC with the proposed multi-task learning (with cross-entropy and contrastive losses) versus matching-only schemes including (Perek et al., 2018). Best results per network are in bold.

et al. (2020b) and Dhungel et al. (2017) obtaining respectively 0.94@0.22 and 0.95@5, while additionally providing accurate dual-view mass correspondences.

Evaluation is supplemented with qualitative results on full mammograms (Fig.4). The additional classification stage (b) helps in eliminating most of false YOLO detections (a). The improvement reached by the combined model (c) compared to the classification-only scheme (b) is highlighted with further wrong proposal removals. For instance, in Fig.4 (2), the number of false positive detections decreased from 7 to 1 from (a) to (b) and further decreased to 0 without any false negatives. In addition, the combined model (c) also successfully identifies the matching patches in both views, which can provide clinicians with reference to further rule out false positives that are difficult to detect, as in Fig.4 (4). Fig.4 demonstrates that variable mass sizes and shapes can be correctly managed. All these findings suggest that exploiting multi-view relationships and multi-tasking learning can greatly guide mammogram interpretation, towards better breast cancer diagnosis and management.

Closer to clinical screening conditions. To evaluate the incidence of false positives under closer to clinical breast screening conditions, a test set of normal mammograms (without masses) has been considered to evaluate our method. Among the 410 INbreast mammograms, 60 CC/MLO image pairs that contain no mass were found. Coarse mass detections are firstly performed on these normal images to generate candidate patches. We use YOLOv3 pre-trained on ImageNet and fine-tuned on 1514 DDSM-CBIS images. Then, a small threshold (10^{-3}) is applied on detection probabilities to ensure that enough predictions from YOLOv3 are selected. We finally obtain 646 candidates patches (374 from CC view, 410 from MLO view), labeled as negative. Then, candidate patches from two views of the same patient are given as inputs of the two branches of our Siamese model for mass/non-mass classification and matching.

We finally obtained 56 false positive predictions whereas 590 true negatives were detected. Accordingly, the obtained specificity was 0.9133. Concerning the 56 false positive detections, only 6 pairs were considered as matched pairs. Thus, the other

Methods	TPR@FPavg	dataset (INbreast)
Kozegar et al. (2013)	0.87 @ 3.67	107
Akselrod-Ballin et al. (2017)	0.93 @ 0.56	100
Ribli et al. (2018)	0.90 @ 0.3	107
Dhungel et al. (2017)	0.95 @ 5	410
Agarwal et al. (2019)	0.92 @ 0.5	107
Yan et al. (2020b)	0.94 @ 0.22	107
Agarwal et al. (2019)	0.74 @ 0.99	70 (inference only)
YOLOv3 only	0.86 @ 1.41	70 (inference only)
classification-only VGG16	0.89 @ 0.29	70 (inference only)
CMCNet VGG16 (ours)	0.96 @ 0.23	70 (inference only)

Table 6. Final detection performance comparisons on INbreast (Moreira et al., 2012) between the proposed method (CMCNet with VGG16 and contrastive loss) and state-of-the-art approaches.

44 patches can be further eliminated because there is no corresponding detection in the other view. These results confirm that our contributions can also provide reliable detection results in a setup more similar to a screening process in real life.

5. Conclusion

In this paper, we propose a novel multi-tasking approach that combines breast mass/non-mass classification with dual-view mass matching between complementary CC/MLO mammograms. We prove the effectiveness of integrating multi-view information within the breast mass detection pipeline by extensive experiments on public datasets. Based on Siamese networks and contrastive learning, our method generalizes well using different deep networks and shows impressive results as an integrated CAD system. We can thus easily address the problem of false detections without struggling with difficult whole-image detection schemes. More globally, the proposed contributions pave the way for robust automatic second opinions in breast cancer diagnosis.

Even if multiple masses can still be detected using the classification network, dealing with more than one mass with respect to matching purposes should deserve further investigation. In future works, our framework could also be extended by associating detection and matching with segmentation techniques to further guide clinicians in their interpretation tasks. Furthermore, it is essential to push further data fusion by extracting and integrating both multi-view and longitudinal information.

Acknowledgements

This work was partly funded by France Life Imaging (grant ANR-11-INBS-0006 from the French *Investissements d'Avenir* program). Authors would like to thank M. Cozic from Medecom (`http://www.medecom.fr/`) for fruitful discussions.

References

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

Akselrod-Ballin, A., Karlinsky, L., Hazan, A., Bakalo, R., Horesh, A.B., Shoshan, Y., Barkan, E., 2017. Deep learning for automatic detection of abnormal findings in breast mammography, in: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support.

Amit, G., Hashoul, S., Kisilev, P., Ophir, B., Walach, E., Zlotnick, A., 2015. Automatic dual-view mass detection in full-field digital mammograms, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 44–52.

Arevalo, J., González, F.A., Ramos-Pollán, R., Oliveira, J.L., Lopez, M.A.G., 2015. Convolutional neural networks for mammography mass lesion classification, in: IEEE Engineering in Medicine and Biology Society.

Carneiro, G., Nascimento, J., Bradley, A.P., 2015. Unregistered multiview mammogram analysis with pre-trained deep learning models, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 652–660.

Choukroun, Y., Bakalo, R., Ben-Ari, R., Akselrod-Ballin, A., Barkan, E., Kisilev, P., 2017. Mammogram classification and abnormality detection from nonlocal labels using deep multiple instance neural network, in: Eurographics Workshop on Visual Computing for Biology and Medicine.

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

Geras, K.J., Wolfson, S., Shen, Y., Wu, N., Kim, S., Kim, E., Heacock, L., Parikh, U., Moy, L., Cho, K., 2017. High-resolution breast cancer screening with multi-view deep convolutional neural networks. arXiv preprint arXiv:1703.07047 .

Hadsell, R., Chopra, S., LeCun, Y., 2006. Dimensionality reduction by learning an invariant mapping, in: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1735–1742.

Hamidinekoo, A., Denton, E., Rampun, A., Honnor, K., Zwiggelaar, R., 2018. Deep learning in mammography and breast histology, an overview and future trends. Medical Image Analysis 47, 45–67.

Han, X., Leung, T., Jia, Y., Sukthankar, R., Berg, A.C., 2015. MatchNet: Unifying feature and metric learning for patch-based matching, in: IEEE Conference on Computer Vision and Pattern Recognition, pp. 3279–3286.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition, in: IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778.

Jørgensen, K.J., Bewley, S., 2015. Breast cancer screening viewpoint of the IARC working group. The New England Journal of Medicine .

Jouirou, A., Baâzaoui, A., Barhoumi, W., 2019. Multi-view information fusion in mammograms: A comprehensive overview. Information Fusion 52.

Koch, G., Zemel, R., Salakhutdinov, R., 2015. Siamese neural networks for one-shot image recognition, in: ICML Deep Learning Workshop.

Kozegar, E., Soryani, M., Minaei, B., Domingues, I., et al., 2013. Assessment of a novel mass detection algorithm in mammograms. Journal of Cancer Research and Therapeutics 9, 592.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks, in: Advances in Neural Information Processing Systems, pp. 1097–1105.

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

Lehman, C.D., Wellman, R.D., Buist, D.S., Kerlikowske, K., Tosteson, A.N., Miglioretti, D.L., 2015. Diagnostic accuracy of digital screening mammography with and without computer-aided detection. Journal of the American Medical Association: Internal Medicine 175, 1828–1837.

Lévy, D., Jain, A., 2016. Breast mass classification from mammograms using deep convolutional neural networks. arXiv preprint arXiv:1612.00542 .

Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., Van Der Laak, J.A., Van Ginneken, B., Sánchez, C.I., 2017. A survey on deep learning in medical image analysis. Medical Image Analysis 42, 60–88.

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.Y., Berg, A.C., 2016. SSD: Single shot multibox detector, in: European Conference on Computer Vision, pp. 21–37.

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

Moreira, I.C., Amaral, I.F., Domingues, I., Cardoso, A.J.M., Cardoso, M.J., Cardoso, J.S., 2012. Inbreast: toward a full-field digital mammographic database. Academic Radiology .

Myers, E.R., Moorman, P., Gierisch, J.M., Havrilesky, L.J., Grimm, L.J., Ghate, S., Davidson, B., Mongtomery, R.C., Crowley, M.J., McCrory, D.C., et al., 2015. Benefits and harms of breast cancer screening: a systematic review. Journal of the American Medical Association 314, 1615–1634.

Perek, S., Hazan, A., Barkan, E., Akselrod-Ballin, A., 2018. Mammography dual view mass correspondence. arXiv preprint arXiv:1807.00637 .

Rakhlin, A., Shvets, A., Iglovikov, V., Kalinin, A.A., 2018. Deep convolutional neural networks for breast cancer histology image analysis, in: International Conference Image Analysis and Recognition, pp. 737–744.

Redmon, J., Farhadi, A., 2018. YOLOv3: an incremental improvement. arXiv preprint arXiv:1804.02767 .

Ren, S., He, K., Girshick, R., Sun, J., 2015. Faster R-CNN: towards real-time object detection with region proposal networks, in: Advances in Neural Information Processing systems, pp. 91–99.

Ribli, D., Horváth, A., Unger, Z., Pollner, P., Csabai, I., 2018. Detecting and classifying lesions in mammograms with deep learning. Scientific Reports 8, 1–7.

Ruder, S., 2017. An overview of multi-task learning in deep neural networks. arXiv preprint arXiv:1706.05098 .

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., et al., 2015. Imagenet large scale visual recognition challenge. International Journal of Computer Vision 115, 211–252.

Shen, L., Margolies, L.R., Rothstein, J.H., Fluder, E., McBride, R., Sieh, W., 2019. Deep learning to improve breast cancer detection on screening mammography. Scientific Reports 9, 1–12.

Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 .

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., 2015. Going deeper with convolutions, in: IEEE Conference on Computer Vision and Pattern Recognition.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z., 2016. Rethinking the inception architecture for computer vision, in: IEEE Conference on Computer Vision and Pattern Recognition, pp. 2818–2826.

Tan, M., Le, Q.V., 2019. EfficientNet: rethinking model scaling for convolutional neural networks. arXiv preprint arXiv:1905.11946 .

Torre, L.A., Islami, F., Siegel, R.L., Ward, E.M., Jemal, A., 2017. Global cancer in women: burden and trends.

Vijayarajan, S., Jaganathan, P., 2014. Breast cancer segmentation and detection using multi-view mammogram. Academic Journal of Cancer Research 7.

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

Warren, R.M., Duffy, S., Bashir, S., 1996. The value of the second view in screening mammography. The British Journal of Radiology 69, 105–108.

Yan, Y., Conze, P.H., Decencière, E., Lamard, M., Quellec, G., Cochener, B., Coatrieux, G., 2019. Cascaded multi-scale convolutional encoder-decoders for breast mass segmentation in high-resolution mammograms, in: International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 6738–6741.

Yan, Y., Conze, P.H., Lamard, M., Quellec, G., Cochener, B., Coatrieux, G., 2020a. Multi-tasking siamese networks for breast mass detection using dual-view mammogram matching, in: International Workshop on Machine Learning in Medical Imaging.

Yan, Y., Conze, P.H., Quellec, G., Lamard, M., Cochener, B., Coatrieux, G., 2020b. Two-stage breast mass detection and segmentation system towards automated high-resolution full mammogram analysis. arXiv preprint arXiv:2002.12079 .

Zagoruyko, S., Komodakis, N., 2015. Learning to compare image patches via convolutional neural networks, in: IEEE Conference on Computer Vision and Pattern Recognition, pp. 4353–4361.

Zhang, X., Zhang, Y., Han, E.Y., Jacobs, N., Han, Q., Wang, X., Liu, J., 2018. Classification of whole mammogram and tomosynthesis images using deep convolutional neural networks. IEEE Transactions on Nanobioscience 17, 237–242.

Zhou, H., Zaninovich, Y., Gregory, C., 2017. Mammogram classification using convolutional neural networks, in: International Conference on Technology Trends.

Zhu, W., Lou, Q., Vang, Y.S., Xie, X., 2017. Deep multi-instance networks with sparse label assignment for whole mammogram classification, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 603–611.