Follow the Radiologist: Clinically Relevant Multi-View Cues for Breast Cancer Detection from Mammograms

Kshitiz Jain[1], Krithika Rangarajan[1,2], and Chetan Arora[1⋆]

[1] Indian Institute of Technology Delhi, Delhi, India
[2] All India Institute of Medical Sciences Delhi, India

Abstract. Automated breast cancer detection using deep learning based object detection models have achieved high sensitivity, but often struggles with high false positive rate. While radiologists possess the ability to analyze and identify malignant masses in mammograms using multiple views, it poses a challenge for deep learning based models. Inspired by how object appearance behaves across multiple views in natural images, researchers have proposed several techniques to exploit geometric correspondence between location of a tumor in multiple views and reduce false positives. We question the clinical relevance of such cues. We show that there is inherent ambiguity in geometric correspondence between the two mammography views, because of which accurate geometric alignment is not possible. Instead, we propose to match morphological cues between the two views. Harnessing recent advances for object detection approaches in computer vision, we adapt a state-of-the-art transformer architecture to use proposed morphological cues. We claim that proposed cues are more agreeable with a clinician's approach compared to the geometrical alignment. Using our approach, we show a significant improvement of 5% in sensitivity at 0.3 False Positives per Image (FPI) on benchmark INBreast dataset. We also report an improvement of 2% and 1% on AIIMS and benchmark DDSM dataset respectively. Realizing lack of open source code base in this area impeding reproducible research, we are publicly releasing source code and pretrained models for this work. [3]

1 Introduction

Screening mammography. Breast cancer is one of the few cancers amenable to screening through Mammography. Mammograms, serving as 2D representations of a 3D anatomical structure (Fig. 1), are acquired in two views: cranio-caudal (CC) and medio-lateral oblique (MLO). It is imperative in clinical practice to see both the views of a breast. A suspicious structure is considered a mass only if it is seen in both the views, if it is not, it is likely to be only an overlap of fibro-glandular tissues due to breast compression during image acquisition.

⋆ Corresponding author
[3] https://mammo-iitd-aiims.github.io/CEN

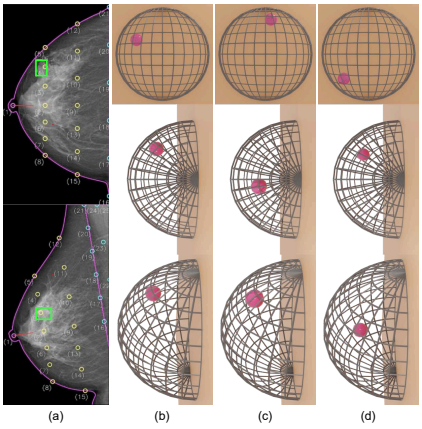

Fig. 1. [Geometric inconsistency]: (a) shows approach by [12,13] which uses mammogram landmarks and cross-view correspondences to localize cancer (green box). (b), (c), and (d), represent 3D breast models with tumors. Top, middle, and bottom row show front, CC and MLO view respectively. Notably, (b) mimics (a), while (c) and (d) reveal significant variations in tumor location across different views, emphasizing limited relevance of geometrical correspondences in reconciling tumor positions.

Deep learning for mammography. Modern deep learning techniques have significantly advanced automated mammogram triage, improving radiologists' detection capabilities during screening [13,12,33,14,19]. Ribli *et al.* [20] employed the Faster R-CNN model for detecting cancer lesions, while Tang *et al.* [25] utilized text embeddings from RoBERTa [11] as a weak supervision technique for detection. Addressing specific challenges in mammogram analysis, Rangarajan *et al.* [16] focused on dense mass detection. In their work [17], they addressed the detection of small mass cancers, contributing to early diagnosis. However, achieving high sensitivity at a low false positive rate remains a challenge.

Multiview approach with global fusion. Previous approaches have explored merging global image-level features from ipsilateral views (CC and MLO) to enhance breast cancer detection precision [23,29,26,24,34,13,33,14,12,19]. However, limited improvement has been observed due to the relatively large size of whole mammograms compared to the region of interest (ROI).

Focusing on ROI with geometric alignment. Recent techniques advocate focusing on specific ROIs relying on geometric relationship between views. Liu *et al.* proposed [13,12], using pseudo landmarks on mammograms and employing bipartite graph convolution to learn geometric and visual features. Yang *et al.* [33] used distance of cancer from the nipple area. Wang *et al.* [28] used asymmetry in bilateral views, and Ma *et al.* [14] introduced Cross-View Relation Block structure to combine geometric and visual features from the two views.

Our claim: geometric alignment is irrelevant. The main thesis of this work is that geometric alignment cues as used in multi-view works currently are irrelevant. We show this visually in Fig. 1. Geometrical Alignment assumes that the breast is a rigid 3-Dimensional structure projected onto a 2D plane. However in reality the breast is a pliable organ, and significant differences in mammography can be generated simply by different levels of compression or rolling the breast differently for compression during the acquisition of mammograms. More details regarding how a radiologist interprets mammograms can be found in Chapter 13: Interpreting the Mammograms of the textbook [7]. Looking at the problem from

the perspective of 3D reconstruction in computer vision, one can find the 3D location of a mass by triangulation, only if a correspondence is known. On the other hand, the geometry cues can at best be used to constrain the correspondence using epipolar lines, but can not be used to establish a correspondence. Therefore, marker based registration (Fig. 1) as suggested in [12][13] or distance from nipple area [32][33] is neither a valid correspondence criterion from computer vision perspective, nor clinically relevant from a radiologist's perspective. Ren *et al.* [19] proposed a detection framework that shares similarities with our work. The authors suggest to refine proposals from a Faster-RCNN model using a greedy approach for proposal matching. However, their method assigns a positive label to both FP-FP (false positive) and TP-TP (true positive) proposal pairs from different views. We note that FP-FP matching could be incorrect, and unlikely to share visual or geometric features. In contrast, we make a stronger claim, that any strict geometrical correspondence between the two views (and not merely FP-FP in [19]) is flawed.

Our Proposal. Acknowledging the limitations of geometric alignment, we propose a novel approach that matches morphological features between ROIs. This allows to use ROIs from other view to expand the contextual information without requiring geometric correspondence. This makes our technique highly interpretable to the clinicians, and also acceptable to the computer vision community.

Contributions. (1) We use a novel framework to model the relationship between two views in mammography which integrates morphological cues from the ROIs in the other view similar to what a radiologist does. **(2)** The proposed approach can be easily integrated with any new deep learning detection models, making it adaptable to future advancements in the field. We harness the pluggable design to propose a transformer-based architecture for breast cancer detection. **(3)** Significantly advancing the state-of-the-art (SOTA) in detection accuracy, our model achieves a sensitivity of 0.55 at 0.1 FPI on DDSM dataset [8] compared to 0.47 by the current SOTA. On our AIIMS dataset we outperformed SOTA by 1.5 times in false positive rate. **(4)** Addressing limitations in existing research and committing to open science principles, we will release our source-code and pre-trained models for transparent and reproducible research.

2 Methodology

Generating proposals. We use state-of-the-art FocalNet-DINO [31] module (referred to as FND hereon), pretrained on natural images, and fine-tune it for breast cancer detection from a single view. We use I_m and I_c to denote original Medio-Lateral Oblique (MLO) and Cranio-Caudal (CC) view mammograms respectively. We pass I_m and I_c separately into FND and generate several candidate predictions indicating the abnormality in that region. We use $P_m = \{P_m^i\}_{i=1}^m$, and $P_c = \{P_c^j\}_{j=1}^n$ to denote these predictions of FND for MLO and CC views respectively. Here, m and n denote the number of predicted boxes by the FND for MLO and CC view respectively. Each of the P_m^i (or P_c^j) contains details of the predicted box B_m^i, and a confidence score c_m^i. Each B_m^i in

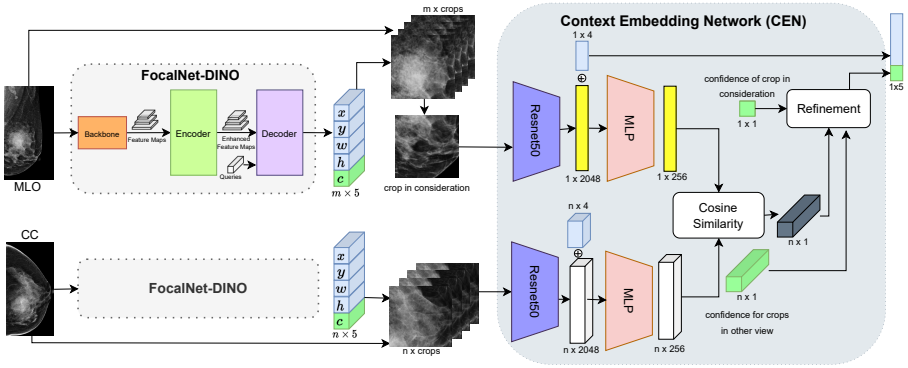

Fig. 2. Model Overview: This figure illustrates the architecture of our proposed model for multi-view breast cancer detection. The model takes mammogram images from two different views as input. The images are processed by an object detection model to generate initial proposals. These proposals are then refined using a proposal refining network, which predicts the correlation between the contextual information of proposal pairs from different views.

turn contains $\{x, y, w, h\}$ denoting the center $(x,\ y)$, width, and height of the predicted box. Whereas c_m^i is a scalar denoting the confidence of the malignancy in the box region. A confidence of zero indicates a benign abnormality.

Context Embedding Network (CEN). As a radiologist would have done, we take each prediction from one view (say MLO view for example), and look for its support from candidate predictions in the other view (CC in this example). Mathematically, for each proposal in one view, we compute an MCS score with every other proposal from the other view, leading to $M_{m\times n} = [m_{ij}]$, where m_{ij} denotes the MCS score between P_m^i and P_c^j. We propose a Siamese Network [1] to compute the MCS score between the proposals. We crop each of the proposal regions from the corresponding mammogram, and pass it through a ResNet50 [5] backbone to extract discriminative features from the proposals. To initialize the weights of ResNet50, we pretrain the model by performing classification on the bounding boxes predicted from the training set. We concatenate the feature embedding from the ResNet50 model with the bounding box information. We pass the concatenated feature vector from three-layer Multi-Layer Perceptron (MLP) with 1024, 512, and 256 nodes in each layer, respectively. ReLU activation functions are applied to the first two layers, resulting in a 256 dimensional embedding vector for each proposal. We use Z_m^i to denote the embedding vector for i^{th} proposal in the MLO view, and similarly Z_c^j for the CC view.

Computing MCS score. We use cosine similarity between feature embeddings of the proposals from two different views. The similarity is computed as:

$$m_{ij} = \frac{Z_m^i \cdot Z_c^j}{|Z_m^i|\ |Z_c^j|}. \tag{1}$$

Confidence update. After computing the MCS scores between each pair of proposals from two views, we update the confidence scores as follows:

$$\tilde{c}_m^i = c_m^i + \max_j \left(m_{ij}\ c_c^j\right), \quad \text{and} \quad \tilde{c}_c^j = c_c^j + \max_i \left(m_{ij}\ c_m^i\right). \tag{2}$$

The expression essentially means that for a proposal in MLO view we look for the support in all the CC proposal (indicated by m_{ij}). The support is modulated by the confidence of the CC proposal itself, implying that we are only looking for proposals indicating malignancy. Recall that no support for malignancy automatically implies negative support as the confidence threshold of the second stage is set higher. We then compute the maximum support among the CC view proposals, and add it to the confidence of the proposal from the MLO view. The operation is symmetrically done for the proposals from CC view as well. Note that while ResNet embedding captures visual information, concatenation of box information allows to use size of the mass also for contextual comparison.

Interpretability of the predictions from proposed model. It may be noted that similar to radiologists, we do not look for a specific proposal in the other view, which may have required geometric correspondence. Instead we are looking for any proposal in the other view which support the current observation morphologically in the view being processed. Hence, we simply take maximum of the MCS scores, and add it to the confidence of current candidate proposal. As we will show later in the experiments section, this keeps the proposed model highly interpretable to the clinicians, as it is possible to visualize which region in the first view our model is attending to, and which region in the other view supported the observation. Further mimicking the clinical process, if there is no supporting evidence in the other view, a Radiologist considers it as a negative support for the malignancy. Similarly, in our model as well, this leads to a negative maximum MCS score, or minimal change in the current malignancy confidence for the region. We keep a lower confidence threshold for detection in the 1st stage, and higher in the 2nd stage. Hence, no support (lower MCS score) from the other view ends up pushing the region below the 2nd stage threshold, making it equivalent to a negative support. Fig. 2 gives the pictorial overview.

Implementation Details. The CEN is trained end-to-end using the cross entropy loss function for each predicted proposal. The proposed framework doesn't attempt to change the bounding box coordinates of a proposal, and hence no regression loss is used. The learning rate is set to 1e-05, and we train the network for 100 epochs on a server with 8 NVidia V100, 32GB GPUs.

3 Results and Discussions

Datasets. We use following datasets in our experiments: **(1) DDSM-CBIS [8]:** The public dataset contains 1324 mammography image pair, of which we use 1054 pairs for training (472 malignant, 582 benign) and 270 for testing (109 malignant, 161 benign). **(2) INBreast [15]:** Consists of 200 image pairs from 115 patients, of which 41 are malignant, and 159 benign. **(3) AIIMS Dataset:**

It is a private diagnostic dataset containing 3,501 mammograms, from 1580 pairs (284 malignant, 1296 benign). We used 445 pairs for training (183 malignant, 262 benign) and 1135 pairs for testing (101 malignant, 1034 benign).

Table 1. Comparison with previous state-of-the-art breast cancer detection approaches. Please refer to main text for details.

Method	Venue	Sensitivity@(their FPI,our FPI)		
Camp. et at.[2]	PMB'04	80@(1.10,**0.33**)	-	-
Eltonsy et at [4]	TMI'07	92@(5.40,**1.17**)	88@(2.40,**0.68**)	81@(0.60,**0.36**)
Sampat et at. [21]	MedPhy'08	88@(2.70,**0.68**)	85@(1.50,**0.47**)	80@(1.00,**0.33**)
CVR-RCNN [14]	ICPR'21	92@(4.40,**1.17**)	88@(1.90,**0.68**)	85@(1.20,**0.47**)
BG-RCNN [13]	CVPR'20	95@(4.40,**1.96**)	92@(1.90,**1.17**)	89@(1.20,**0.73**)
MommiNet-v2 [33]	MIA'21	90@(2.00,**0.87**)	85@(1.00,**0.47**)	80@(0.50,**0.33**)
AG-RCNN [12]	TPAMI'21	96@(4.40,**2.39**)	92@(1.90,**1.17**)	90@(1.20,**0.87**)

Evaluation metric. We use widely reported Free-Response Receiver Operating Characteristic (FROC) curve [3] for our experiments. The curve provides a graphical representation of sensitivity/recall values at different false positives per image (FPI). We follow related works in this area [16], and consider a prediction as true positive if center of the predicted bounding box lies within the ground-truth bounding box. Other than detection region, we also evaluate the performance of proposed model for classification task. We report results in terms of Accuracy, F1-Score, Precision, Recall, and AUC score. We classify an image as malignant if there is at least one bounding box with malignant prediction.

3.1 Comparison with SOTA breast cancer detection methods

AG-RCNN [12] is an existing benchmark work on mammogram detection but with different experiment settings, we follow their methodology to split the DDSM dataset for training, validation, and testing set. Results are shown in Tab. 1 where we compare the FROC metric for various approaches using single-view [2,4,21], multi-view using ipsilateral views [14,12,13] and tri-view using both ipsilateral and bilateral views[33]. These are the only methods to our knowledge reporting results on the DDSM dataset. Rest of the methods have reported results on private datasets and do not release their model/code, making it infeasible to compare. To effectively evaluate with prior methods, we keep the same recall/sensitivity value and compare the FPI values (lower is better). From Tab. 1 it is evident that our method achieves significantly low FPI at same recall, and hence outperforms all the previous state-of-the-art methods.

Previous methods often rely on subsets of public datasets to demonstrate their research findings. For instance, Yan et al. [30] reported a sensitivity of 0.96 at 0.23 FPI for the INBreast dataset, while our approach achieves a sensitivity of

Table 2. Detection (recall at different FPI) and classification comparison results. Please refer to main text for details.

			Detection Results						Classification Results		
	Model Name	**Venue**	**R@0.025**	**R@0.05**	**R@0.1**	**R@0.3**	**R@0.5**	**R@1**	**Accuracy**	**F1-Score**	**AUC-Score**
DDSM	Faster RCNN[18]	TPAMI'17	0.02	0.04	0.07	0.13	0.22	0.29	0.620	0.294	0.604
	DN DEF.[9]	CVPR'22	0.32	0.37	0.40	0.46	0.48	0.53	0.756	0.675	0.803
	YOLO-V8[6]	Ultralytics'23	0.11	0.14	0.20	0.32	0.36	0.42	0.644	0.394	0.601
	DINO [35]	ICLR'23	0.22	0.27	0.39	0.55	0.62	0.68	0.757	0.644	0.799
	FND[31]	NeurIPS'22	0.29	0.35	0.47	0.70	0.76	**0.86**	0.733	0.654	0.789
	FND + OURS		**0.36**	**0.45**	**0.55**	**0.71**	**0.77**	0.84	**0.761**	**0.676**	**0.809**
AIIMS	Faster RCNN[18]	TPAMI'17	0.20	0.26	0.32	0.47	0.53	0.66	0.920	0.288	0.761
	DN DEF.[9]	CVPR'22	0.64	0.72	0.74	0.78	0.81	0.82	0.952	0.680	0.949
	YOLO-V8[6]	Ultralytics'23	0.19	0.27	0.38	0.51	0.61	0.71	0.916	0.212	0.779
	DINO[35]	ICLR'23	0.13	0.33	0.41	0.48	0.61	0.91	0.851	0.260	0.649
	FND [31]	NeurIPS'22	0.64	0.78	0.84	0.91	0.93	**0.97**	0.949	0.706	0.964
	FND + OURS		**0.70**	**0.80**	**0.86**	**0.93**	**0.95**	0.96	**0.958**	**0.747**	**0.976**
INBreast	Faster RCNN[18]	TPAMI'17	0.16	0.19	0.25	0.40	0.45	0.54	0.828	0.297	0.661
	DN Def[9]	CVPR'22	0.17	0.26	0.39	0.44	0.47	0.51	0.825	0.526	0.772
	YOLO-V8[6]	Ultralytics'23	0.02	0.03	0.06	0.08	0.10	0.12	0.258	0.346	0.541
	DINO[35]	ICLR'23	0.20	0.30	0.35	0.54	0.65	0.82	0.820	0.275	0.738
	FND [31]	NeurIPS'22	0.28	**0.46**	0.56	0.73	0.75	0.82	0.865	0.641	0.869
	FND + OURS		**0.34**	0.42	**0.63**	**0.78**	**0.82**	**0.84**	**0.872**	**0.666**	**0.887**

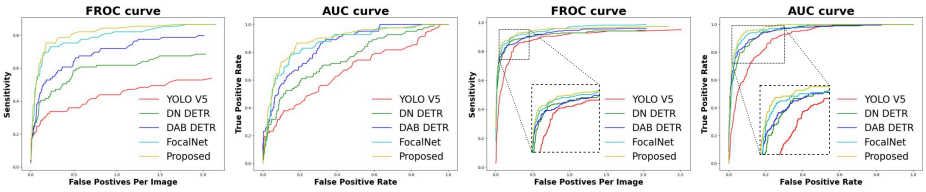

Fig. 3. First and second images show FROC and AUC curves respectively on the INBreast dataset. Third and fourth images show the same on AIIMS dataset.

0.94 at merely 0.025 FPI, representing a tenfold improvement, in one of our cross-validation splits. Similarly, Chu et al. [27] reported Sensitivity@FPI values of 87.5@0.5, 91.5@1.0, and 94@2.0 on their DDSM data split. Our method surpasses this baseline, achieving values of 92.2@0.5, 94.6@1.0, and 97.5@2.0 on our custom split. However, we note that a fair comparison is challenging, as both references did not use standard test data, and dataset split details are undisclosed.

3.2 Comparison with natural image based techniques

Quantitative results. Tab. 2 shows results from various techniques on the three datasets. For AIIMS dataset, we use a 0.25:0.03:0.72 train-val-test split. For DDSM we use the standard split as described earlier. The results show that FDN itself improves most of the SOTA techniques. Our CEN module further improves the FDN results, particularly at low FPI rates. For INBreast, since the dataset is small, it is not possible to train a model using only the INBreast. Hence, we conduct two experiments. In Tab. 2 we show the results when using AIIMS dataset for training and whole INBreast dataset for inference. In Tab. 3, we show results of training on AIIMS dataset, and 5-fold cross-validation on

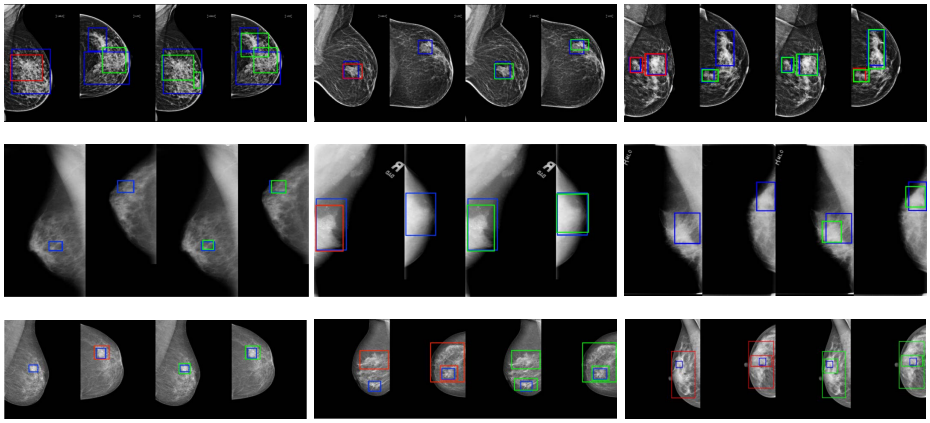

Fig. 4. Illustrative results of the proposed system on diverse datasets (1st row: private dataset, 2nd row: DDSM, 3rd row: INBreast). Ground truth bounding boxes are depicted in blue, while red bounding boxes represent predictions from the FND model and green bounding boxes represent predictions from the proposed method. Each subfigure comprises four mammography images, including a set of MLO and CC views. The first pair showcases the current SOTA(FND) predictions, while the second pair displays the predictions of the proposed model. Notably, the proposed model leverages contextual cues from the complementary view, leading to improved predictions.

Table 3. Cross-validation results on INBreast, showing mean and standard deviation in sensitivity values at different FPI.

FPI	**0.025**	**0.050**	**0.100**	**0.300**	**1.000**
FND [31]	0.534(±0.250)	0.630(±0.192)	0.745(±0.109)	0.827(±0.114)	0.874(±0.110)
FND+OURS	**0.601**(±0.230)	**0.651**(±0.190)	**0.807**(±0.080)	**0.873**(±0.089)	**0.927**(±0.041)

INBreast(finetuning on 4-folds, and testing on 5th fold). We report average and standard deviation of sensitivity values obtained at various FPI values.

Classification Performance. In addition to evaluating for detecting cancerous regions (region predictions are important to ensure interpretability), we also evaluated our model for image classification task. Results, AUC plots, and visualizations, are shown in Tab. 2, Fig. 3 and Fig. 4 respectively.

Ablation study. Supplementary material contains results of following ablation studies: **(1)** Effect of refinement module on detected proposals, **(2)** Integration with baselines[9,10,6] other than FND, **(3)** Using other texture cues[22] in CEN, **(4)** Using only proposal size and removing location information in the CEN, and **(5)** Using attention-based context modeling in the CEN.

4 Conclusion

Current techniques for multiview breast cancer detection, either use full image fusion, which is not effective, or use geometric alignment which is clinically irrel-

evant. We propose a new framework for fusing ipsilateral mammography views based on morphological features from ROIs. This allows to focus on interpretable ROIs, yet do not use problematic geometric alignment. Our framework can work with multiple deep neural detection backbones, and achieves significant improvements in recall, particularly at low false positives per image (FPI) levels.

Acknowledgments. We thank Mr. Rohan Raju Dhanakshirur and Mr. Devesh Pant for their assistance. We acknowledge and thank the funding support from Department of Biotechnology, India vide grant number BT/PR33193/AI/133/5/2019, and AIIMS Delhi-IIT Delhi Center of Excellence in AI funded by Ministry of education, government of India, Central Project Management Unit, IIT Jammu with sanction number IITJMU/CPMU-AI/2024/0002. Kshitiz Jain is supported by Yardi School of Artificial Intelligence, IIT Delhi via its Publication Grant for Students.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Bromley, J., Guyon, I., LeCun, Y., Säckinger, E., Shah, R.: Signature verification using a siamese time delay neural network. NeurIPS (1993)
2. Campanini, R., Dongiovanni, D., Iampieri, E., Lanconelli, N., Masotti, M., Palermo, G., Riccardi, A., Roffilli, M.: A novel featureless approach to mass detection in digital mammograms based on support vector machines. Physics in Medicine & Biology (2004)
3. Egan, J.P., Greenberg, G.Z., Schulman, A.I.: Operating characteristics, signal detectability, and the method of free response. The Journal of the Acoustical Society of America (1961)
4. Eltonsy, N.H., Tourassi, G.D., Elmaghraby, A.S.: A concentric morphology model for the detection of masses in mammography. IEEE TMI (2007)
5. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. CVPR (2016)
6. Jocher, G., Chaurasia, A., Qiu, J.: YOLO by Ultralytics (Jan 2023), https://github.com/ultralytics/ultralytics
7. Kopans, D.B.: Breast imaging. Lippincott Williams & Wilkins (2007)
8. Lee, R.S., Gimenez, F., Hoogi, A., Miyake, K.K., Gorovoy, M., Rubin, D.L.: A curated mammography data set for use in computer-aided detection and diagnosis research. Nature Scientific data (2017)
9. Li, F., Zhang, H., Liu, S., Guo, J., Ni, L.M., Zhang, L.: DN-DETR: Accelerate DETR training by introducing query denoising. CVPR (2022)
10. Liu, S., Li, F., Zhang, H., Yang, X., Qi, X., Su, H., Zhu, J., Zhang, L.: DAB-DETR: Dynamic anchor boxes are better queries for DETR. arXiv:2201.12329 (2022)
11. Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., Stoyanov, V.: Roberta: A robustly optimized bert pretraining approach. arXiv:1907.11692 (2019)
12. Liu, Y., Zhang, F., Chen, C., Wang, S., Wang, Y., Yu, Y.: Act like a radiologist: towards reliable multi-view correspondence reasoning for mammogram mass detection. IEEE TPAMI (2021)

13. Liu, Y., Zhang, F., Zhang, Q., Wang, S., Wang, Y., Yu, Y.: Cross-view correspondence reasoning based on bipartite graph convolutional network for mammogram mass detection. CVPR (2020)
14. Ma, J., Li, X., Li, H., Wang, R., Menze, B., Zheng, W.S.: Cross-view relation networks for mammogram mass detection. ICPR (2021)
15. Moreira, I.C., Amaral, I., Domingues, I., Cardoso, A., Cardoso, M.J., Cardoso, J.S.: INBreast: toward a full-field digital mammographic database. Academic radiology (2012)
16. Rangarajan, K., Agarwal, P., Gupta, D.K., Dhanakshirur, R., Baby, A., Pal, C., Gupta, A.K., Hari, S., Banerjee, S., Arora, C.: Deep learning for detection of isodense, obscure masses in mammographically dense breasts. European Radiology (2023)
17. Rangarajan, K., Gupta, A., Dasgupta, S., Marri, U., Gupta, A.K., Hari, S., Banerjee, S., Arora, C.: Ultra-high resolution, multi-scale, context-aware approach for detection of small cancers on mammography. Nature Scientific Reports (2022)
18. Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: Towards real-time object detection with region proposal networks. IEEE TPAMI (2017)
19. Ren, Y., Lu, J., Liang, Z., Grimm, L.J., Kim, C., Taylor-Cho, M., Yoon, S., Marks, J.R., Lo, J.Y.: Retina-match: ipsilateral mammography lesion matching in a single shot detection pipeline. MICCAI (2021)
20. Ribli, D., Horváth, A., Unger, Z., Pollner, P., Csabai, I.: Detecting and classifying lesions in mammograms with deep learning. Nature Scientific reports (2018)
21. Sampat, M.P., Bovik, A.C., Whitman, G.J., Markey, M.K.: A model-based framework for the detection of spiculated masses on mammography. Medical physics (2008)
22. Singh, H., Sharma, V., Singh, D.: Comparative analysis of proficiencies of various textures and geometric features in breast mass classification using k-nearest neighbor. Visual Computing for Industry, Biomedicine, and Art (2022)
23. Sun, L., Wang, J., Hu, Z., Xu, Y., Cui, Z.: Multi-view convolutional neural networks for mammographic image classification. IEEE Access (2019)
24. Sun, Z., Jiang, H., Ma, L., Yu, Z., Xu, H.: Transformer based multi-view network for mammographic image classification. MICCAI (2022)
25. Tang, Y., Cao, Z., Zhang, Y., Yang, Z., Ji, Z., Wang, Y., Han, M., Ma, J., Xiao, J., Chang, P.: Leveraging large-scale weakly labeled data for semi-supervised mass detection in mammograms. CVPR (2021)
26. van Tulder, G., Tong, Y., Marchiori, E.: Multi-view analysis of unregistered medical images using cross-view transformers. MICCAI (2021)
27. Vu, Y.N.T., Guo, D., Taha, A., Su, J., Matthews, T.P.: M&m: Tackling false positives in mammography with a multi-view and multi-instance learning sparse detector. arXiv:2308.06420 (2023)
28. Wang, X., Tan, T., Gao, Y., Han, L., Zhang, T., Lu, C., Beets-Tan, R., Su, R., Mann, R.: Disasymnet: Disentanglement of asymmetrical abnormality on bilateral mammograms using self-adversarial learning. MICCAI (2023)
29. Wu, N., Phang, J., Park, J., Shen, Y., Huang, Z., Zorin, M., Jastrzebski, S., Févry, T., Katsnelson, J., Kim, E., et al.: Deep neural networks improve radiologists' performance in breast cancer screening. IEEE TMI (2019)
30. Yan, Y., Conze, P.H., Lamard, M., Quellec, G., Cochener, B., Coatrieux, G.: Towards improved breast mass detection using dual-view mammogram matching. MedIA (2021)
31. Yang, J., Li, C., Dai, X., Gao, J.: Focal modulation networks. NeurIPS (2022)

32. Yang, Z., Cao, Z., Zhang, Y., Han, M., Xiao, J., Huang, L., Wu, S., Ma, J., Chang, P.: MommiNet: Mammographic multi-view mass identification networks. MICCAI (2020)
33. Yang, Z., Cao, Z., Zhang, Y., Tang, Y., Lin, X., Ouyang, R., Wu, M., Han, M., Xiao, J., Huang, L., et al.: Momminet-v2: Mammographic multi-view mass identification networks. MedIA (2021)
34. You, K., Lee, S., Jo, K., Park, E., Kooi, T., Nam, H.: Intra-class contrastive learning improves computer aided diagnosis of breast cancer in mammography. MICCAI (2022)
35. Zhang, H., Li, F., Liu, S., Zhang, L., Su, H., Zhu, J., Ni, L., Shum, H.Y.: DINO: DETR with improved denoising anchor boxes for end-to-end object detection. ICLR (2023)