Glandular Distribution–Aware Pretraining for Architectural Distortion Detection in Digital Breast Tomosynthesis

Xiaoling Liu[1,†], Zilong He[2,†], Wei Liang[1,3,†], Yue Li[4], Yun Teng[5], Chushan Lin[2], Jiahui Jia[2], Qianli Zhong[2], Yile Bin[2], Jiayi Li[2], Xiangyuan Ma[1,*], Weiguo Chen[2,*], Yao Lu[6,*]

[1] Department of Biomedical Engineering, College of Engineering, Shantou University, Shantou, China
[2] Department of Diagnostic Imaging, Nanfang Hospital, Southern Medical University, Guangzhou, China
[3] Department of Computer Science & Engineering, Lehigh University, Bethlehem, PA, USA
[4] Department of Radiology, University of Pennsylvania, Philadelphia, PA, USA
[5] Department of Radiology, Lianjiang People's Hospital, Lianjiang, China
[6] School of Computer Science and Engineering, Sun Yat-sen University, Guangzhou, China

2025 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) | 979-8-3315-1557-7/25/$31.00 ©2025 IEEE | DOI: 10.1109/BIBM66473.2025.11356891

***Abstract*—Architectural distortion (AD) is a subtle early sign of breast cancer in digital breast tomosynthesis (DBT). Numerous computer-aided detection methods have been developed to assist radiologists in identifying AD. Despite advances, existing approaches diverge from clinical practice by relying heavily on annotated abnormalities while neglecting abundant normal DBT volumes that are critical for understanding glandular distributions and detecting abnormalities. Motivated by this clinical insight, we propose a Glandular Distribution-Aware pretraining paradigm that learns intrinsic glandular spatial priors from unlabeled normal DBT scans in a self-supervised manner, which are then transferred to downstream AD detection. Specifically, we design a masked inpainting GAN to pretrain an encoder that captures both local tissue continuity and global anatomical context. This encoder is then fused with a COCO-pretrained ResNet backbone within a RetinaNet detector, coupling general object-detection capabilities with glandular-specific sensitivity. Extensive experiments on an in-house DBT dataset demonstrate that our method achieves a mean true positive fraction of 64.51% and a sensitivity of 75.18% at 2.0 false positives per volume, significantly outperforming state-of-the-art methods and better aligning with clinical radiology practice.**

***Index Terms*—Architectural Distortion, Digital Breast Tomosynthesis, Self-Supervised Pretraining, Glandular Distribution, Computer-Aided Detection**

I. Introduction

Breast cancer remains a major global public health challenge, and early detection through breast cancer screening is essential for improving treatment outcomes [1], [2]. During screening, architectural distortion (AD) is recognized as one of the typical presentations of breast cancer [3].

Over the past decade, digital breast tomosynthesis (DBT) has been widely adopted, offering higher sensitivity and lower false-positive rates [4]. Despite advances, the volumetric nature of DBT data significantly prolongs radiologist interpretation time. Consequently, computer-aided detection (CADe) systems are being developed to automate image analysis and streamline clinical workflow [5], [6].

In recent years, numerous CADe systems have been proposed to localize AD in DBT [7]–[13]. Despite these advances, methods still depend on limited annotated volumes while disregarding the vast majority of normal examinations. This diverges from screening scenarios, where normal examinations far outnumber abnormal examinations, which account for less than 1% (1,081 of 121,652) [14]. Naturally, normal examinations will serve as critical baselines for radiologists to learn glandular distribution and detect abnormalities [15]. In healthy breasts, mammary ducts and lobules exhibit a characteristic orientation converging to the nipple [16], whereas AD manifests as disruption of this organized pattern [17]. Thus, exposed to unlabeled normal volumes, models enable to learn the distribution of healthy glandular tissue, improving sensitivity to AD.

Therefore, in scenarios with limited annotated lesions, emulating radiologists' reliance on unlabeled normal volumes is critical to enhance glandular distribution feature learning for improved detection. To achieve this, it is necessary to develop methods that can effectively leverage this abundant, unlabeled normal image.

Significant progress in self-supervised learning for unlabeled images has been underpinned by the development of Generative Adversarial Networks (GANs) [18]. Extending the GAN-based framework to breast imaging, Konz et al. [19] and Swiecicki et al. [20] applied inpainting-based unsupervised learning to normal DBT, achieving anomaly localization by measuring reconstruction discrepancies. However, confined to healthy volumes during training, these models failed to acquire lesion-specific features, yielding heatmaps instead of precise localization and classification.

Motivated by the aforementioned GAN-based inpainting framework and the reliance of radiologists on normal volumes, we propose a Glandular Distribution–Aware (GDA) Pretraining strategy. This approach leverages abundant unlabeled DBT

† Xiaoling Liu, Zilong He and Wei Liang contributed equally to this work
* To whom correspondence should be addressed: maxiangyuan@stu.edu.cn,chen1999@smu.edu.cn, and luyao23@mail.sysu.edu.cn

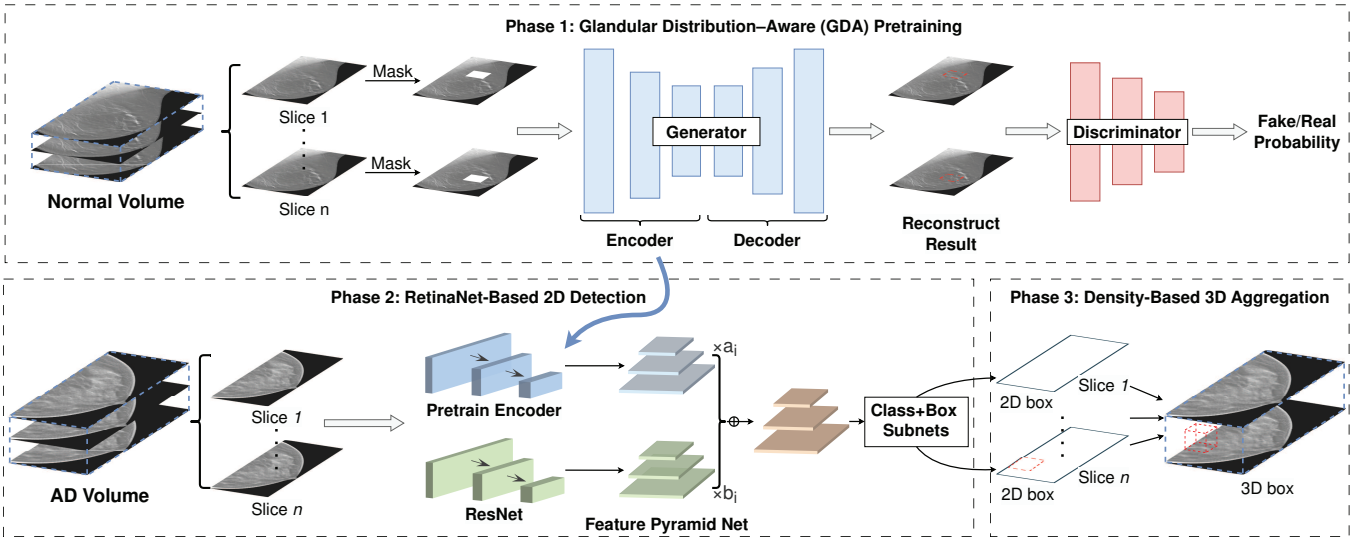

Fig. 1: Overview of the proposed AD detection framework: Phase 1 employs GAN-based inpainting on unlabeled normal volumes to learn glandular distribution features and pretrains an encoder; Phase 2 integrates this encoder into a RetinaNet-based backbone to process 2D slices and generates bounding boxes with classification confidences; Phase 3 then aggregates 2D candidates via a density-based clustering algorithm to assemble the final 3D candidates.

volumes in a self-supervised manner to learn features of glandular distribution and further improve supervised AD detection. In summary, the contributions of this paper consist of three aspects:

- We propose a pretraining–transfer paradigm tailored for screening scenarios with scarce abnormal examples. Leveraging abundant normal DBT volumes, our self-supervised pretraining learns glandular distribution representations that serve as strong priors for downstream AD detection, as AD typically manifests as disruptions of these normal patterns.
- To effectively learn healthy glandular distribution priors, we design a GDA masked-inpainting GAN pretraining strategy that reconstructs occluded regions from context under the guidance of a multi-scale discriminator enforcing both local and global anatomical consistency.
- To enhance the downstream RetinaNet detector, we integrate the task-specific glandular distribution priors with general object-detection capabilities through a parallel architecture comprising a GDA encoder and a COCO-pretrained ResNet. Ablation studies demonstrate that both branches are critical for optimal detection performance.

II. Method

This section outlines the proposed AD detection pipeline, as illustrated in Fig. 1, consisting of three sequential phases: (1) GDA Pretraining; (2) RetinaNet-Based 2D Detection; and (3) Density-Based 3D Aggregation. Phase 1 leverages GAN-based inpainting on unlabeled normal volumes for glandular-aware self-supervised pretraining, yielding a feature-enriched encoder. Phase 2 embeds this encoder into RetinaNet for slice-wise lesion detection. Phase 3 clusters the 2D detections through a density-based algorithm to generate final 3D candidates.

A. Glandular Distribution–Aware (GDA) Pretraining

As showed in Fig. 1 Phase 1, we introduce the glandular distribution-aware, self-supervised pretraining framework built upon a masked-inpainting GAN [21], comprising a generator and a discriminator. The generator receives normal, unlabeled DBT slices with randomly masked regions and attempts to reconstruct the missing content using the surrounding context. Then the discriminator learns to distinguish between original and reconstructed patches. Through adversarial training, the generator's encoder learns latent representations that capture glandular patterns by jointly optimizing reconstruction and adversarial objectives.

1) Generator: U-Net architecture [22] serves as our generator since it has been widely validated in medical imaging [23]. It takes an input image with a randomly occluded 128×128 region and reconstructs the missing content solely from the surrounding context. The 128×128 mask size follows prior experimental validation [19], [20].

2) Discriminator: The discriminator is trained on real and synthetic DBT patches to learn to distinguish between them. It adopts a multi-scale Pix2pixHD architecture, inspired by Jiang et al. [24] and Wang et al. [25]. For each generated masked sample, the enclosing 256×256 contextual patch is randomly extracted. This entire patch, containing the inpainted region and surrounding native tissue, is fed into the discriminator. By processing both the mask and its context (128/256 nesting), the discriminator jointly evaluates local realism and global anatomical continuity, consistent with previous empirical findings [19], [20].

B. RetinaNet-based 2D Detection

RetinaNet serves as a strong single-stage baseline with demonstrated cross-task generalization [26]. We preserve these

TABLE I: Performance on comparison study

Model	R@0.5(%)	R@1.0(%)	R@2.0(%)	R@4.0(%)	FPs@0.8	MTPF(%)	*p*-value
RetinaNet	56.51	66.46	71.30	75.06	8.0857	61.42	<0.05
Faster RCNN	55.53	66.46	73.83	78.26	6.6219	61.36	<0.05
YOLOv12	50.86	61.67	70.02	74.20	—	57.21	$\ll$0.01
DETR	43.37	59.95	68.30	71.38	5.8303	54.35	$\ll$0.01
Proposed	**59.95**	**68.67**	**75.18**	**79.85**	**4.0667**	**64.51**	—

R@k represents sensitivity R at k false positives per image; '-' represents sensitivity less than 80%; and the bold value represents the best results in each comparison experiment.

TABLE II: Performance on ablation study

Model	R@0.5(%)	R@1.0(%)	R@2.0(%)	R@4.0(%)	FPs@0.8	MTPF(%)	*p*-value
Baseline	56.51	66.46	71.30	75.06	8.0857	61.42	<0.05
GDA Only	44.10	53.81	64.00	72.97	6.5788	50.94	<0.05
Fused (Proposed)	**59.95**	**68.67**	**75.18**	**79.85**	**4.0667**	**64.51**	—

R@k represents sensitivity R at k false positives per image; and the bold value represents the best results in each comparison experiment.

default architectural settings and, on this foundation, integrate the pretrained GDA encoder as a parallel backbone to the COCO-pretrained ResNet. This configuration couples the COCO-pretrained backbone's general object-detection representational capacity with task-specific sensitivity to glandular distribution cues. As illustrated in Fig. 1 Phase 2, each backbone independently extracts features that are processed by its respective FPN to generate multi-scale representations. Finally, at each pyramid level i, the corresponding feature maps from the two branches are added using learnable scalar weights a_i and b_i.

C. Density-Based 3D Aggregation

Since DBT images are volumetric, We follow the density–based method [10]–[12] to aggregate 2D detection candidates across slices to produce final 3D results.

III. Experiments and Results

A. Dataset and Preprocessing

Two distinct DBT datasets were employed: one for self-supervised pretraining on normal breast anatomy and another for supervised detection of AD.

1) Pretraining Dataset: For self-supervised pretraining, we utilized a dataset of DBT images gathered from the Duke Health System [9], which contained 19,737 normal reconstruction volumes from 4,897 screening examinations of 4,609 normal participants with a mean age of 44-66 years.

2) Detection Dataset: With institutional review board approval, we curated an annotated in-house DBT dataset from Nanfang Hospital, Southern Medical University, Guangzhou, China, comprising 1214 volumes of 500 participants aged 20–78 years. All images were captured using the Selenia Dimensions Mammography System by Hologic (1 mm interval, 0.0866–0.1093 mm pixel spacing).

B. Evaluation

The optimal model checkpoints are selected and evaluated for their performance using two metrics: the mean true positive fraction (MTPF), calculated at a range of 0.05–2.0 false positives per volume (FPs), and the sensitivity at 2.0 FPs, derived from the free-response receiver operating characteristic (FROC) [5]. To assess the statistical significance of performance differences between methods, we computed p-values using bootstrap analysis with 1000 patient-based resampled datasets derived from the test results.

C. Comparison with the Competing Methods

The proposed detection method was evaluated against four representative baseline models: anchor-based detectors RetinaNet [26] and Faster R-CNN [13], anchor-free detector YOLOv12 [27], and the transformer-based model DETR [28]. Table I reveals that the proposed method significantly outperforms competing approaches ($p < 0.05$).

D. Ablation Study

We conducted an ablation study to isolate the impact of our GDA encoder and its integration strategy. Three variants of the RetinaNet detector were evaluated:

1) Baseline: Standard RetinaNet with a COCO-pretrained ResNet backbone.
2) GDA Only: RetinaNet with only the GDA-pretrained backbone.
3) Fused (Proposed): RetinaNet with parallel COCO-ResNet and GDA-pretrained backbones.

Table II shows that using only the GDA encoder lowered performance, suggesting task-agnostic pretraining alone is insufficient. In contrast, the fused configuration improved MTPF from 61.42% to 64.51%, confirming the benefit of coupling general object-detection representational capacity with task-specific sensitivity.

IV. Conclusions

In this paper, we propose a self-supervised pretraining strategy that leverages abundant unlabeled normal DBT volumes to pretrain an encoder for capturing glandular distribution features in AD detection, emulating radiologists' practice.

This encoder is then fused with a COCO-pretrained ResNet backbone within a RetinaNet detector, coupling general object-detection capacity with glandular-specific sensitivity. Extensive experiments on in-house DBT datasets demonstrate the effectiveness of our method, achieving an MTPF of 64.51% and 75.18% sensitivity at 2.0 false positives per volume, which outperforms other detectors.

ACKNOWLEDGMENT

This work was supported in part by Guangdong Basic and Applied Basic Research Foundation under Grant 2023A1515011488; in part by Young Scientists Fund of the National Natural Science Foundation of China, 82502292; in part by the Guangzhou Municipal Science and Technology Project under Grant 2025A04J4114; in part by the Natural Science Foundation of Guangdong Province under Grant 2414050003969; in part by the National Key&D Program of China under Grant 2023YFE0204300; in part by the R&D project of Pazhou Lab (HuangPu) under Grant 2023K0606; in part by the National Natural Science Foundation of China under Grants 82441027 and 62371476; in part by the Guangzhou Science and Technology Bureau under Grant 2023B03J1237; in part by the Health Research Major Projects of Hunan Health Commission under Grant W20241010; and in part by the Guangdong Province Key Laboratory of Computational Science at the Sun Yat-sen University under Grant 2020B1212060032.

REFERENCES

[1] A. Q. Khan, M. Touseeq, S. Rehman, M. Tahir, M. Ashfaq, E. Jaffar, and S. F. Abbasi, "Advances in breast cancer diagnosis: a comprehensive review of imaging, biosensors, and emerging wearable technologies," *Frontiers in Oncology*, vol. Volume 15 - 2025, 2025.

[2] A. N. Giaquinto, H. Sung, L. A. Newman, R. A. Freedman, R. A. Smith, J. Star, A. Jemal, and R. L. Siegel, "Breast cancer statistics 2024," *CA: a cancer journal for clinicians*, vol. 74, no. 6, pp. 477–495, 2024.

[3] S. Banik, R. M. Rangayyan, and J. L. Desautels, "Detection of architectural distortion in prior mammograms," *IEEE transactions on medical imaging*, vol. 30, no. 2, pp. 279–294, 2010.

[4] T. Li, Y.-R. Su, J. M. Lee, E. S. O'Meara, D. L. Miglioretti, K. Kerlikowske, L. Henderson, and N. Houssami, "Tomosynthesis vs digital mammography screening in women with a family history of breast cancer," *JAMA oncology*, 2025.

[5] J. Pan, Z. He, Y. Li, W. Zeng, Y. Guo, L. Jia, H. Jiang, W. Chen, and Y. Lu, "Atypical architectural distortion detection in digital breast tomosynthesis: a multi-view computer-aided detection model with ipsilateral learning," *Physics in Medicine & Biology*, vol. 68, no. 23, p. 235006, 2023.

[6] J. Bai, R. Posner, T. Wang, C. Yang, and S. Nabavi, "Applying deep learning in digital breast tomosynthesis for automatic breast cancer detection: A review," *Medical image analysis*, vol. 71, p. 102049, 2021.

[7] G. Palma, S. Muller, I. Bloch, and R. Iordache, "Detection of convergence areas in digital breast tomosynthesis using a contrario modeling," in *Medical Imaging 2009: Computer-Aided Diagnosis*, vol. 7260. SPIE, 2009, pp. 217–224.

[8] G. Palma, I. Bloch, and S. Muller, "Detection of masses and architectural distortions in digital breast tomosynthesis images using fuzzy and a contrario approaches," *Pattern Recognition*, vol. 47, no. 7, pp. 2467–2480, 2014.

[9] M. Buda, A. Saha, R. Walsh, S. Ghate, N. Li, A. Święcicki, J. Y. Lo, and M. A. Mazurowski, "A data set and deep learning algorithm for the detection of masses and architectural distortions in digital breast tomosynthesis images," *JAMA network open*, vol. 4, no. 8, pp. e2 119 100–e2 119 100, 2021.

[10] Y. Li, Z. He, Y. Lu, X. Ma, Y. Guo, Z. Xie, G. Qin, W. Xu, Z. Xu, W. Chen *et al.*, "Deep learning of mammary gland distribution for architectural distortion detection in digital breast tomosynthesis," *Physics in Medicine & Biology*, vol. 66, no. 3, p. 035028, 2021.

[11] Y. Li, Z. He, X. Ma, W. Zeng, J. Liu, W. Xu, Z. Xu, S. Wang, C. Wen, H. Zeng *et al.*, "Architectural distortion detection based on superior–inferior directional context and anatomic prior knowledge in digital breast tomosynthesis," *Medical Physics*, vol. 49, no. 6, pp. 3749–3768, 2022.

[12] Y. Li, Z. He, J. Pan, W. Zeng, J. Liu, Z. Zeng, W. Xu, Z. Xu, S. Wang, C. Wen *et al.*, "Atypical architectural distortion detection in digital breast tomosynthesis: a computer-aided detection model with adaptive receptive field," *Physics in Medicine & Biology*, vol. 68, no. 4, p. 045013, 2023.

[13] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *Advances in neural information processing systems*, vol. 28, 2015.

[14] C. K. Abbey, A. I. Bandos, M. K. Parthasarathy, M. A. Webster, and M. L. Zuley, "Changes in reader performance during sequential reading of breast cancer screening digital breast tomosynthesis examinations," *Radiology*, vol. 313, no. 2, p. e232885, 2024.

[15] C. F. Nodine, H. L. Kundel, S. C. Lauver, and L. C. Toto, "Nature of expertise in searching mammograms for breast masses," *Academic radiology*, vol. 3, no. 12, pp. 1000–1006, 1996.

[16] J. Chakraborty, A. Midya, S. Mukhopadhyay, R. M. Rangayyan, A. Sadhu, V. Singla, N. Khandelwal, P. Bhattacharyya, and P. M. Azevedo-Marques, "Detection of the nipple in mammograms with gabor filters and the radon transform," *Biomedical Signal Processing and Control*, vol. 15, pp. 80–89, 2015.

[17] E. A. Sickles, "Acr bi-rads® atlas, breast imaging reporting and data system." *American College of Radiology.*, p. 39, 2013.

[18] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," *Advances in neural information processing systems*, vol. 27, 2014.

[19] N. Konz, H. Dong, and M. A. Mazurowski, "Unsupervised anomaly localization in high-resolution breast scans using deep pluralistic image completion," *Medical image analysis*, vol. 87, p. 102836, 2023.

[20] A. Swiecicki, N. Konz, M. Buda, and M. Mazurowski, "A generative adversarial network-based abnormality detection using only normal images for model training with application to digital breast tomosynthesis," *Scientific Reports*, vol. 11, 05 2021.

[21] D. Pathak, P. Krahenbuhl, J. Donahue, T. Darrell, and A. A. Efros, "Context encoders: Feature learning by inpainting," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2536–2544.

[22] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*. Springer, 2015, pp. 234–241.

[23] R. Azad, E. K. Aghdam, A. Rauland, Y. Jia, A. H. Avval, A. Bozorgpour, S. Karimijafarbigloo, J. P. Cohen, E. Adeli, and D. Merhof, "Medical image segmentation review: The success of u-net," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024.

[24] G. Jiang, J. Wei, Y. Xu, Z. He, H. Zeng, J. Wu, G. Qin, W. Chen, and Y. Lu, "Synthesis of mammogram from digital breast tomosynthesis using deep convolutional neural network with gradient guided cgans," *IEEE Transactions on Medical Imaging*, vol. 40, no. 8, pp. 2080–2091, 2021.

[25] T.-C. Wang, M.-Y. Liu, J.-Y. Zhu, A. Tao, J. Kautz, and B. Catanzaro, "High-resolution image synthesis and semantic manipulation with conditional gans," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 8798–8807.

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

[27] Y. Tian, Q. Ye, and D. Doermann, "Yolov12: Attention-centric real-time object detectors," *arXiv preprint arXiv:2502.12524*, 2025.

[28] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, "End-to-end object detection with transformers," in *European conference on computer vision*. Springer, 2020, pp. 213–229.