SIFT-DBT: SELF-SUPERVISED INITIALIZATION AND FINE-TUNING FOR IMBALANCED DIGITAL BREAST TOMOSYNTHESIS IMAGE CLASSIFICATION

Yuexi Du[1], Regina J. Hooley[2], John Lewin[2], Nicha C. Dvornek[1,2]

Yale SCHOOL OF ENGINEERING & APPLIED SCIENCE

[1]Department of Biomedical Engineering, Yale School of Engineering & Applied Science, New Haven, CT, USA
[2]Department of Radiology & Biomedical Imaging, Yale School of Medicine, New Haven, CT, USA

Background & Motivation

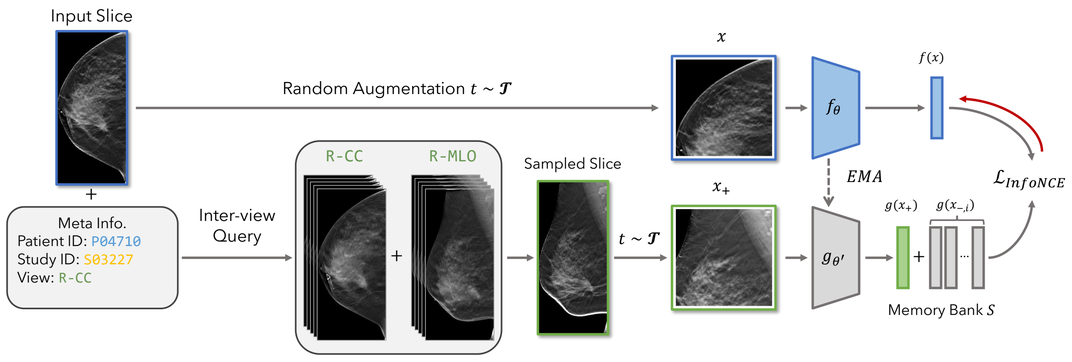
- **Breast cancer** is the most commonly diagnosed cancer globally and the leading cause of cancer-related mortality in women.
- **3D Digital Breast Tomosynthesis (DBT)** has emerged as a powerful imaging tool in breast cancer detection. Different from traditional 2D mammography, DBT provides a more comprehensive z-dimensional view of breast tissue by rotating the X-ray generator around the breast to reconstruct the 3D information, resulting in significantly enhanced resolution of tissue details.
- However, DBT greatly increases the amount of digital imaging data, which further amplifies the challenge of **data imbalance**.
- In one of the largest public DBT datasets BCS-DBT [1], only **101 studies (224 volumes, ~1%)** are labeled as abnormal among a total number of **4,838 studies (19,148 volumes, ~99%)**.
- Among each single volume, only **~2%** area is labeled as ROI, indicating tumor/suspicious region.
- Trivial deep learning classification system fails due to extreme data imbalance issue.

More time spent on filtering out healthy scans!

Our Goal

- Build a fast deep learning **DBT classification system** that can accurately classify *Normal* and *Abnormal* DBT slices & volumes.
- Address the problem of **extreme real-world data imbalance** issue while maintaining robust performance.
- Focus on both local and global high-resolution details via multi-patch finetuning and multi-instance decision making

Method: Self-supervised Initialization (SI)



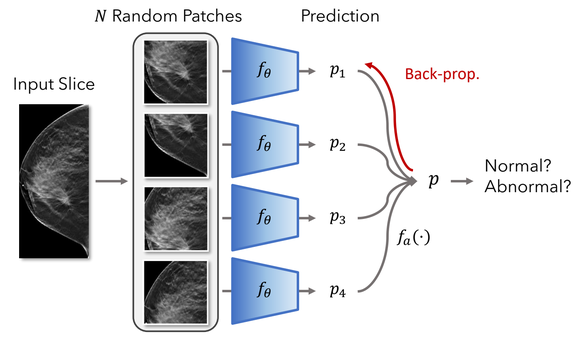
(a) Contrastive Pre-training

- **Contrastive Learning**: To avoid the model collapse issue due to direct training on the imbalanced data, we introduce contrastive learning to force the model to *focus on the local image feature*, regardless of the label. We optimize the classic *InfoNCE* loss here.

$$\mathcal{L}_{InfoNCE} = -\log \frac{\exp(x \cdot x_+/\tau)}{\sum_{i=1}^{S} \exp(x \cdot x_i/\tau)}$$

- **Local Patch Augmentation**: We randomly crop smaller patches (~25% area) to keep high-resolution detail while reducing computational cost. The model also learns to focus on the local features.
- **Inter-slice Positive Sampling**: Inspired by MedAug [2], we sample positive pairs based on meta information of the DBT image. We use random slices from the same patient, same study, different view (CC vs. MLO) as the positive image x_+ to the input x. Imaging of the same breast in a different view should naturally contain related information.
- **Details**: We adapt MoCo-v2 [3] as our base contrastive learning method and sample slices from the other view with 50% probability. We test both ResNet50 and a simple 7 layers CNN model as our backbone model.

Method: Local Multi-Patch Fine-tuning (FT)

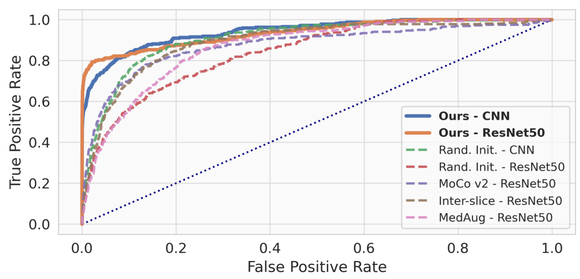


(b) Multi-Patch Fine-tuning & Inference

- **Discriminative Finetuning**: We reduce the learning rate by $\eta = 2.8$ for every layer in the pretrained model to maintain the learned low-level feature extractor while adapting deeper layers to Normal vs. Abnormal classification task.
- **Test-time Multi-patch Prediction**: We use an aggregation function $f_a(\cdot) = avg(\cdot)$ to aggregate prediction results from $N = 20$ randomly cropped sub-patches to make the final slice level predication.
- **Parameter-free Volume Level Prediction**: We use the $maximum$ slice abnormal score as the volume score and adjust the prediction threshold to minimize the gap between normal and abnormal recall, i.e., balanced accuracy prediction, to make the final volume level prediction.

Results: Slice Level

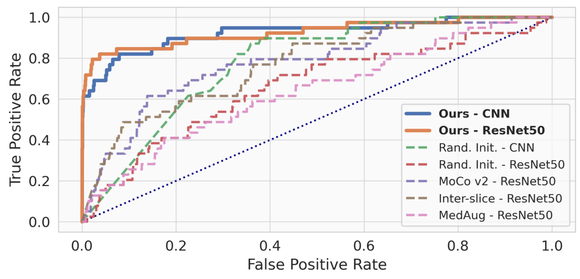
Method	Model	BCS-DBT Slice-level Results					
		AUC	NPV	NR	AR	SP@87	SP@80
Rand. Init.	CNN	87.94	99.91	59.10	**90.31**	71.33	77.23
	ResNet50	85.03	99.76	86.03	61.25	64.22	68.85
MoCo v2	ResNet50	89.80	99.75	93.40	56.56	76.65	83.74
MedAug		86.89	99.86	78.97	79.37	69.80	78.68
Inter-slice		88.67	99.84	87.42	74.69	77.29	84.38
Ours$_{N=1}$	CNN	89.73	99.83	86.36	73.44	76.78	83.62
	ResNet50	90.31	99.89	81.13	83.13	76.69	83.05
Ours	CNN	**94.26**	99.95	**98.98**	60.94	**86.85**	92.98
	ResNet50	93.67	**99.98**	76.38	88.44	82.20	**97.16**



- Classification of test set with **245,875 normal** and **320 abnormal** slices
- We note multiple baselines collapsed with low normal/abnormal recall.
- Our model outperforms all the baselines with better performance.

Results: Volume Level

Method	Model	BCS-DBT Volume-level Results					
		AUC	NPV	NR	AR	SP@87	SP@80
Rand. Init.	CNN	77.79	**100.00**	12.76	**100.00**	64.03	67.43
	ResNet50	65.93	99.38	64.19	61.54	21.63	48.31
MoCo v2	ResNet50	77.89	99.64	65.74	76.92	39.51	64.05
MedAug		63.95	99.30	60.47	58.97	24.87	35.71
Inter-slice		77.23	99.73	48.66	87.18	55.22	58.86
Ours$_{N=1}$	CNN	78.85	99.64	74.73	74.35	52.84	65.85
	ResNet50	69.55	99.67	24.36	92.31	31.05	55.89
Ours	CNN	92.58	99.81	**84.15**	84.62	**82.71**	93.52
	ResNet50	**92.69**	99.81	83.97	84.62	80.86	**97.81**



- Classification of test set with **3,755 normal** and **39 abnormal** volumes.
- Baselines failed by predicting everything as abnormal .
- Our model maintains a stable prediction with higher normal and abnormal recall and also shows a much higher AUC score.

Conclusion

- We proposed a novel slice- and volume-level DBT classification framework called ***SIFT-DBT***, with both **S**elf-supervised pre-trained weight **I**nitialization and multi-patch **F**ine-**T**uning to address the extreme data imbalanced issue for real-world **DBT** data.
- We demonstrate the capability of our model on a DBT test set with 970 patients [1] with a considerable gap comparing with other baselines.
- The proposed method may potentially **speed up** the DBT screening process by automatically filtering out high confidence healthy DBT volumes, allowing the radiologists to focus on high-risk samples first.

References:

1. N.Konz et.al. "*A competition, benchmark, code, and data for using artificial intelligence to detect lesions in digital breast tomosynthesis*." JAMA, 2023
2. Y.N.T. Vu, R et.al. "*Medaug: Contrastive learning leveraging patient metadata improves representations for chest x-ray interpretation*." Machine Learning for Healthcare Conference. PMLR, 2021.
3. K, He et.al. "*Momentum contrast for unsupervised visual representation learning*." CVPR, 2020.

This work was supported by NIH grant R21EB032950.

Contact:
yuexi.du@yale.edu

IEEE INTERNATIONAL SYMPOSIUM ON BIOMEDICAL IMAGING ISBI 2024
27-30 MAY, 2024 – ATHENS, GREECE
MEGARON ATHENS INTERNATIONAL CONFERENCE CENTRE (MAICC)

arXiv

GitHub