

# Contrasting Multiple Mammogram Views for Breast Cancer Classification

Nasif Zaman

*Dept. of Computer Science and Engineering  
University of Nevada, Reno  
Reno, USA  
zaman@nevada.unr.edu*

Alireza Tavakkoli

*Dept. of Computer Science and Engineering  
University of Nevada, Reno  
Reno, USA  
tavakkol@unr.edu*

Md Abu Sayed

*Dept. of Computer Science and Engineering  
University of Nevada, Reno  
Reno, USA  
msayed@nevada.unr.edu*

George Bebis

*Dept. of Computer Science and Engineering  
University of Nevada, Reno  
Reno, USA  
bebis@cse.unr.edu*

**Abstract**—Correspondence finding between different mammogram views is a challenging task. Graph convolutions have proved powerful in mammographic view registration in the past. However, they have been applied exclusively to mass segmentation. In this study we investigate how the same principles can be applied to microcalcification segmentation. However, the results show some requirements need to be fulfilled to encode the exceedingly small microcalcifications.

**Index Terms**—Mammograms, Multi-View Mask RCNN, Graph Convolutional Neural Network, AGN

## I. INTRODUCTION

The Multiview mammogram images provide crucial complementary information about the breast anatomical structures that can be used to interpret digital mammography images for mass detection. In this work, we are considering focusing mainly on the fusion of multiple mammogram views (CC, MLO) for cancer detection in breast mammograms. With the literature review that we have done, we believe we can leverage Liu et al.'s Act Like a Radiologist paper's [1] main ideas (using graph convolution to learn to reason bilateral and ipsilateral views of the breast and detect mass) and add our own enhancements (particular data augmentation schemes to improve the graph convolution and learn to reason temporally to localize changes in breast mass, calcification, and architectural distortion). We will mainly use the CBIS-DDSM dataset [8] for training but we are also considering using Breast Micro-Calcification Dataset [10] for the temporal aspects.

## II. LITERATURE REVIEW

The current literature on mass detection from mammograms can be classified into traditional and deep-learning approaches. Traditional methods rely on handcrafted features of mammograms to detect mass from the image [1]. However, with the revolution of deep learning in the last decade, it has been greatly applied to medical image analysis problems. Researchers developed more accurate mass detection models

with deep convolutional neural networks (CNN) [2]. A typical deep CNN usually tries to reduce the false positives, however, has an inferior performance because of not being end-to-end trainable [1]. Recently, researchers have tried to use state-of-the-art modern object detection algorithms like Faster R-CNN, FPN, Masked R-CNN to tackle this problem for mammogram mass detection. Even though these techniques perform well to some extent, the complementary multi-view mammogram images are not considered in the detection process [6].

Wu et al. [2] propose a two-stage DNN architecture that utilizes multi-view images to train and diagnose with their model. They design a patch-level network to learn pixel-level features and use another network alongside to learn macroscopic breast-level representations. Also, they use a customized ResNet-based network as a backbone of these networks to adapt to high-resolution mammogram images. Another noticeable contribution of their work was pretraining the network on screening BI-RADS classification to handle more noisy labels. In addition, they experimented with four different combinations of views (view-wise, image-wise, breast-wise, joint) as input to the network and combined them to get the results. They trained and evaluated on over 200000 exams (over 1000000 images) with AUC metrics. To corroborate these findings, they conducted a reader study with 14 readers who each reviewed 720 screening mammography tests. When given identical data, they claim that their model is as accurate as professional radiologists.

Ma et al. [6] propose a multi-view framework called Cross-View Relation Region-based Convolutional Neural Networks (CVR-RCNN). They introduce the ipsilateral property to the model and added a relation module to the Faster RCNN that aims to learn ipsilateral inter-proposal relations. As the ipsilateral geometric and semantic relations are not explicitly considered, the relation learning lacks clear constraints. Therefore, the learned relations cannot precisely map between multiple views. Also, the relation module greatly depends on

the quality of the region proposals in the first stage. In addition, the performance of the detection model drops significantly in case of severe gland occlusions.

Due to the strong representation capability of Graph Neural Networks (GCN) for non-Euclidean data and the strong reasoning power of domain knowledge, it has been introduced in visual reasoning tasks such as visual tracking [7]. Li et al. [7] propose a combination of graph convolutional units to learn the graph representation of a 2D image. Liu et al. [1] propose an anatomy aware GCN for mammogram mass detection, capable of detection with multi-view reasoning. They designed their detection and reasoning architecture imitating what a Radiologist might do when they look at mammograms to diagnose. They use CVR-RCNN as a backbone to convert the examined view (the primary view that has best features for mass and calcification), auxiliary view (the other view of the same breast), and contralateral view (the same view as examined but from the other breast) (from CC, MLO) of the patient’s mammogram into a feature space. Then, introduced Bipartite Graph Convolutional Network (BGN) and Inception Graph Convolutional Network (IGN) to learn intrinsic geometric and semantic relation of ipsilateral views, and structural similarities of bilateral view respectively. Both BGN and IGN construct graphs that propagate the multi-view information through nodes. The learned features from examined view together with auxiliary and contralateral views equips the model to visualize interpretable visual cues for clinical diagnosis. They experimented on DDSM and their in-house dataset and evaluated their model performance by false positive per image (FPI), or recall (R) at t. It performs well and provides clear visual cues for the probable lesions.

In [3], Yang et al. propose a model, acronymized as MommiNet-v2, that utilizes DNN based tri-view mass identification approach. They added IpsiDualNet-v2 and BiDualNet-v2 that perform bilateral (main, RMLO (aux)) and ipsilateral (main, “LCC (aux)”) analysis of mammograms view consecutively. They claim that these novel high-resolution networks (HRNet)-based architectures are able to learn the symmetry and geometry constraints. Then, the model aggregates output information from high-resolution representations of all these views and result in improved mass detection performance. They also use a multi-task learning technique to the malignancy data from biopsies and BI-RADS categories, resulting in enhanced mass malignancy classification. Their experiment and performance evaluation procedure is similar to Liu et al. [1]. However, their optimum results are found at R@2.0 instead of Liu’s R@4.0. Also, Liu et al.’s AGN perform slightly better than MommiNet-v2 in most criteria.

### III. METHODOLOGY

Our objective is to fuse multiple mammogram views as [1] have done. The following “Figure. 1” shows (purple boxes) how the authors have incorporated CC, MLO and other breast views to increase the segmentation capability of the system. However, their objectives were limited to mass segmentation only. However, we’ve identified that calcifications are a more

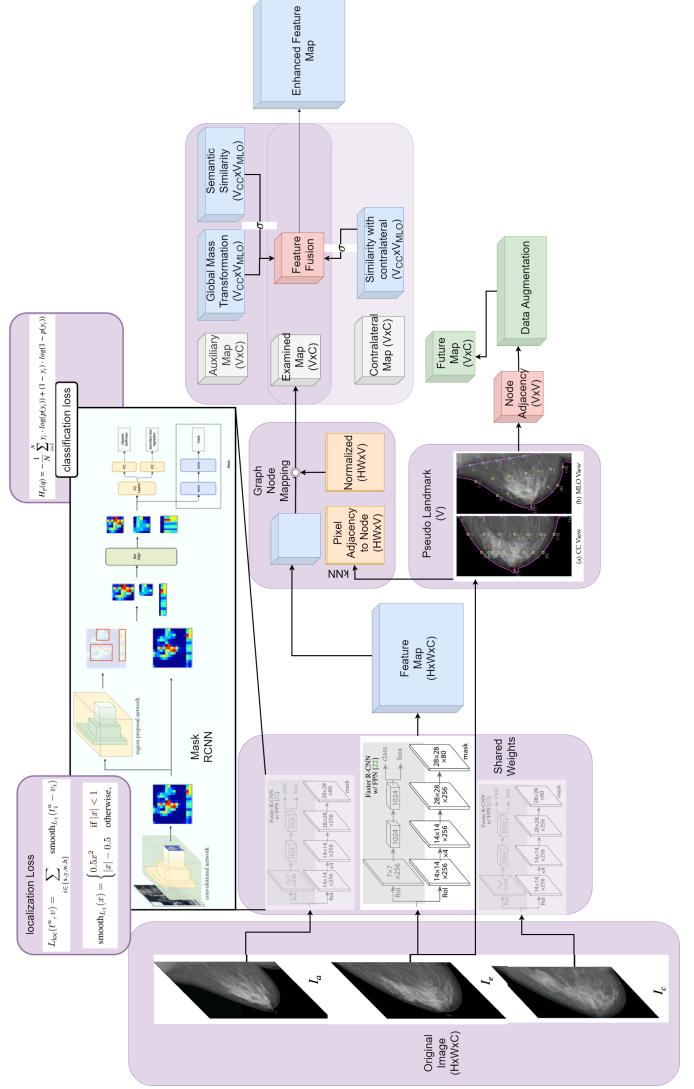


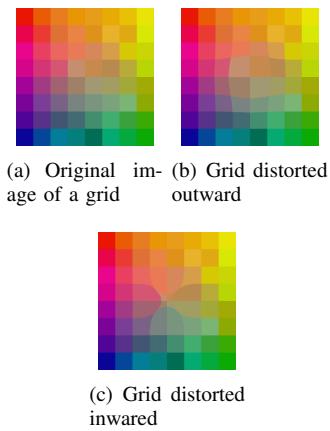
Fig. 1. Architecture overview.

potent use-cases for their graph convolutional approach as the spatial distribution of the calcification nodes can further enhance the geometric learning of multi-view mammogram images. Additionally, based on our understanding of the graph mapping that takes place for examined, auxiliary, and contralateral views, we believe tweaking the node distance value and  $k$  (in kNN) can help create artificial architectural distortions as a data augmentation scheme. In the following figure (green boxes) we show what extensions we have planned and the estimated timeline for the project.

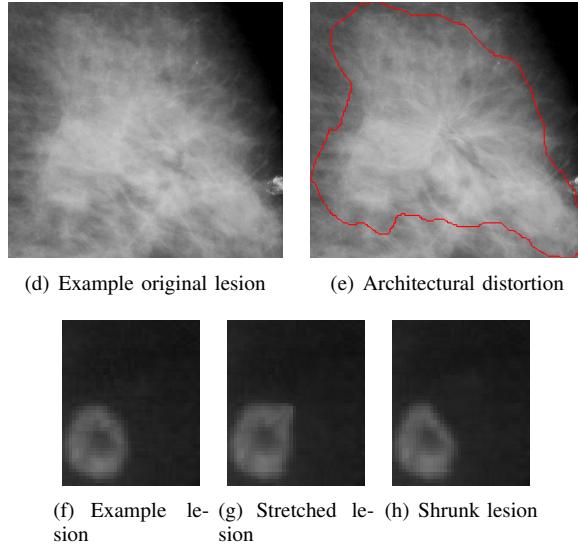
#### A. Preprocessing

Each view of the input images is resized to same size as the examined image in order to maintain the same spatial resolution across different views of the image. Furthermore, simple resizing where aspect ratios are distorted, is unfavorable for image registration. So, we kept the aspect ratio same with padding.

We introduced several architectural distortion with lens distortion effects. The idea is illustrated by the grid images shown below:



In execution, we find that the distortions are not as reminiscent of typical architectural distortion. However, it works quite well in shrinking or stretching compact (non-elongated) masses as illustrated in the following example:

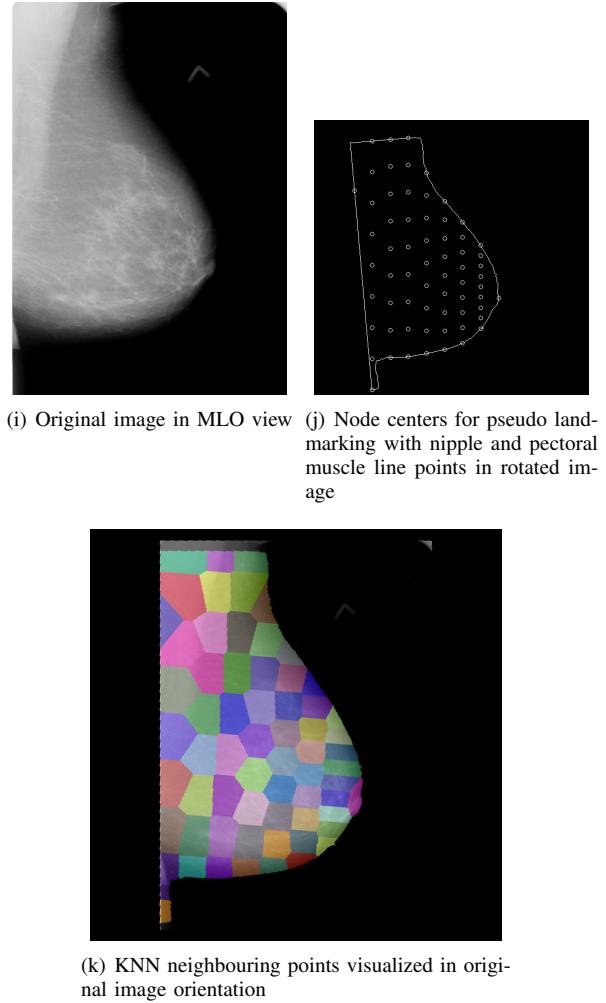


### B. Graph Region Mapping

Our approach to pseudo landmark generation was different to their Hough transform based approach. The primary challenge of pseudo landmark generation is to find the nipple and the pectoral muscle line. That task is fairly straightforward for CC views as Left CC views have the pectoral muscle at the left edge and Right CC view has the pectoral muscle at the right edge.

It gets more challenging in MLO views as the pectoral muscle can be oriented at any position. To obtain the nipple, we first rotated the MLO views until an extremity in the breast contour was found. This extremity was the nipple and the current edge of the rotated image was the pectoral muscle

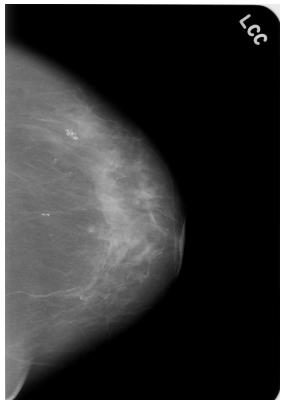
line. This approach produced very good results in obtaining pectoral muscle and nipple:



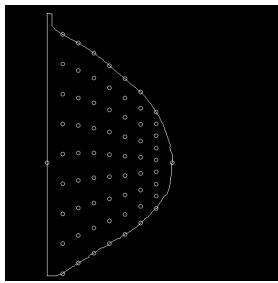
In this phase, feature maps generated from the Mask-RCNN are mapped with the anatomy aware divided regions of the mammogram. Here,  $\Phi$  is a  $V \times C$  dimensional matrix that holds the feature representation of each node.

The next step involved dividing each view into multiple regions or pseudolandmarks so that each region showed some consistent spatial (distance from nipple) or geometric (similar size and shape) relations. For the purposes of our experiment, we divided each CC view into 65 regions and each MLO view into 82 regions. This selection corresponds to the highest performing node numbers in the paper [1].

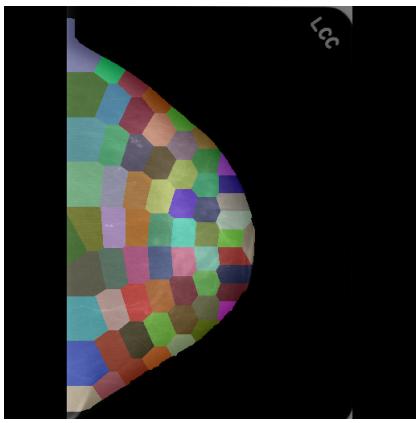
In this study we focused primarily on implementing the cross-view mammogram fusion as outlined in the paper [1]. Our goal was to apply the network on calcification dataset such as the Temporal microcalcification dataset for improved calcification segmentation. The first challenge in applying the same architecture was that no correspondence was established between the numerous calcifications that are present in CC and MLO views. Therefore, we adopted a hough-transform based calcification registration procedure, described below:



(l) Original image in CC view



(m) Node centers for pseudo landmarking with nipple and pectoral muscle line points



(n) KNN neighbouring points visualized

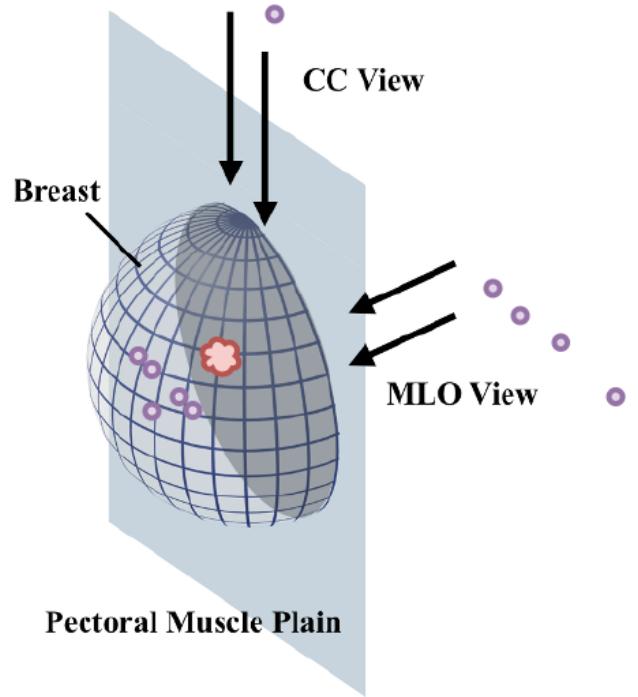


Fig. 2. Geometric Correspondence through spatial voting

#### C. Calcification Correspondence for Geometric Relation Learning

If multiple masses exist on a single mammogram view, the size variation between masses help establish correspondence between views. Specifically in our case, we sorted masses by their sizes for each auxiliary view and assigned the biggest mass in the CC view to the biggest mass in MLO view. However, such correspondence is impossible to establish in microcalcifications as their sizes tend to have very small variance, nullifying this approach.

In figure 2 we can see that a single position in CC view can correspond to multiple positions in the MLO view. Instead of looking throughout the image for geometric correspondence we designed a voting system where only regions of similar distance from the nipple where microcalcifications exist get half votes. It is possible to improve this strategy further by subdividing the views into even more regions, however that comes at a cost of significant performance drop

#### D. Graph Convolution Operations

In [1] the authors establish that graph convolutions can be effective for image registration operation. One intuitive way we can understand the effectiveness of graph convolutions for

establishing correspondence is that while standard 2D convolution operations operate spatially to find semantic relation between pixels, graph convolutions can operate at arbitrary adjacent nodes to find semantic relations. First step towards applying graph convolutions is to create graphs  $\{V, E\}$  or nodes and edges between them. In accordance with the paper, each region in the pseudolandmark subdivision represents a region. Each of these nodes have features. However, it is apparent that each region is not exactly the same size therefore a mechanism is needed to convert them to similarly shaped features for operation. Here, the initial features are the outputs from the backbone or feature pyramid network of the Mask RCNN. The FPN network outputs 5 spatial feature sets with 256 depth. We selected the 4 highest spatial resolutions for graph-based augmentation.

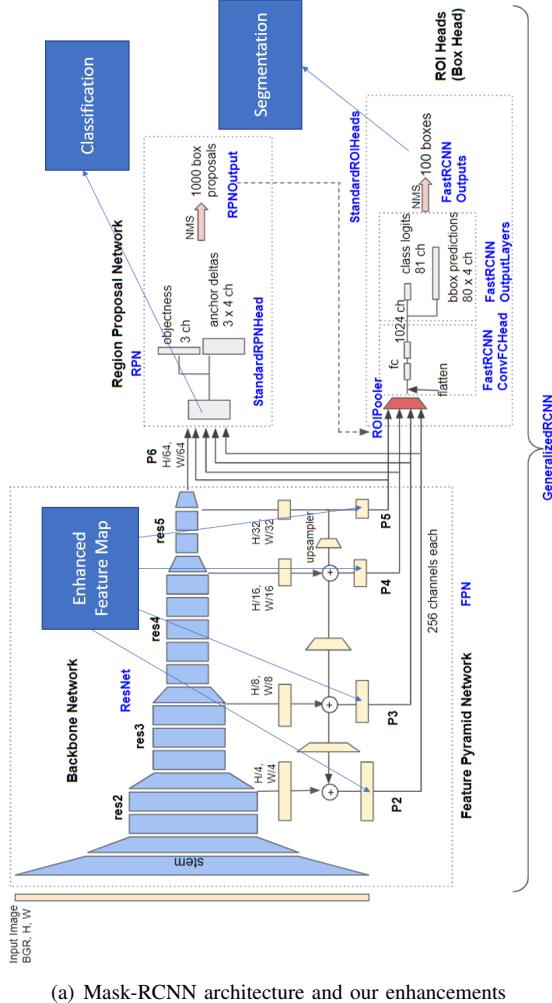
The paper [1] does not mention how many layers of graph convolution they performed, or if they down sampled graph node features or upsampled them. We chose to use two graph convolution operation for correspondence between auxiliary views. The first layer downsampled the node feature length from 256 to 16 and the second layer upsampled it to 256 from 16. Our rational for this approach is to force the network to encode only important features from the two different views.

Likewise, for the ipsilateral views, we first used an encoding graph convolution and secondly used a decoding graph convolution.

#### E. Pathology Classification

In [1], the authors only performed instance segmentation. However, cross-view relations can help determine pathology

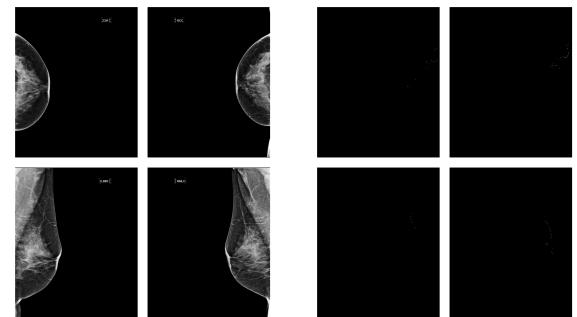
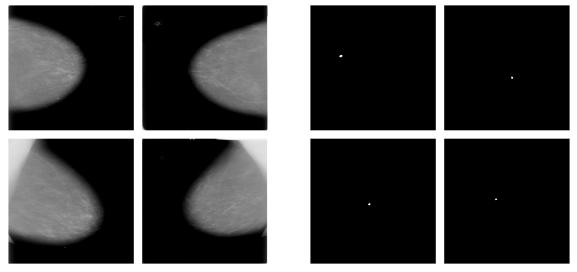
of breasts as well. As bilateral malignancy is highly rare, classifications that identify both breasts as malignant have a high likelihood of being wrong. We take the fused feature representation after the bipartite and ipsilateral graph convolution operations and flatten the feature to perform dense layer operation and classification.



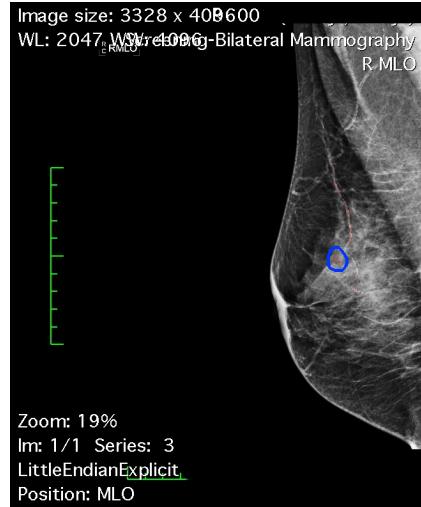
#### IV. DATASETS

Datasets [8] is best suited to train the multi-view and calcification aspects of our objective. This same dataset was used by the authors of [1]. However, while they report that [8] has 1592 cases of mass and 1511 examples of microcalcification of all-view examples, we only found 84 cases with 4 views for a total of 336 images. We split it into train (214), test (68) and validation (54). [10] has temporal microcalcification data with contralateral view supplanted by prior view. [8] train (252), test (80) and validation (64)

For the temporal calcification dataset, contralateral views are not available. However, auxiliary views are available. So, for the time being, we are training by flipping the orientation of the prior image to produce a synthetic contralateral view. Example:



(d) Original images (e) Mask images



In CBIS, the mask images were given as binary files, but in the temporal calcification dataset, the annotated images were color images, so we had to extract the ground truth binary mask files from the colored images. This dataset had 100 cases (100\*4=400 images).

#### A. Code

Our full code is available at <https://github.com/Znasif/ContrastMammogram>. We used Detectron2 (<https://github.com/facebookresearch/detectron2/tree/main/detectron2>) for the Mask RCNN architecture and Pytorch Geometric

TABLE I  
TABLE 1

Method	R@0.5	R@1.0	R@2.0	R@3.0	R@4.0
Our Mask RCNN	59.3	61.3	62.3	63.6	64.1
Our Faster RCNN	58.4	60.2	61.7	63.5	63.8
AG-RCNN	82	89	92.1	93.8	95.5
Our AG RCNN	59.1	62.3	63.5	63.9	64.2

TABLE II  
TABLE 2

Method	R@0.5	R@1.0	R@2.0	R@3.0	R@4.0
Temporal	10.3	19.5	20.6	25.1	26.5

([https://github.com/pyg-team/pytorch\\_geometric](https://github.com/pyg-team/pytorch_geometric)) for graph convolutions.

## V. EXPERIMENTS

We performed the following experiments:

- 1) Segmentation Experiments on only Mask RCNN on CBIS
- 2) Segmentation Experiments with Anatomy-aware Graph RCNN on CBIS
- 3) Classification from Graph Representation
- 4) Segmentation Experiments on Temporal Calcification dataset

We utilize stochastic gradient descent (SGD) for the training with a learning rate 0:02, weight decay 10e-4, momentum 0:9 and nesterov set True. The training process takes 30 epochs in all.

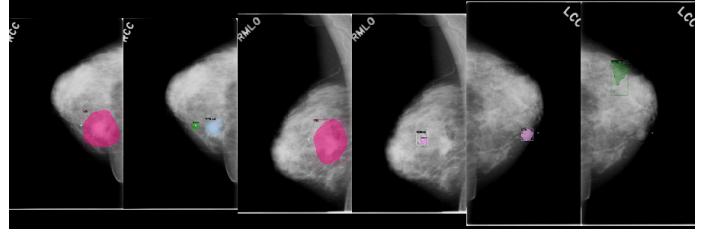
## VI. RESULTS

As a baseline, we trained the Mask RCNN and Faster RCNN without any graph-based augmentation on the CBIS dataset. In the table below we compare the results we got from our system against the ones the authors report:

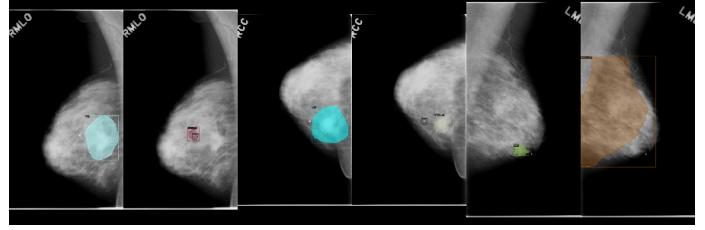
We evaluate the performance by recall (R) at t (t 2 f0:5; 1:0; 2:0; 3:0; 4:0g) false positive per image (FPI), which is simplified as R@t. A mass region is recalled when its IOU (Intersection Over Union) value is larger than 0.2. The reason their models performed better may simply be because of their larger dataset. The reason the temporal dataset was performing so much worse may simply be because of the extremely small size of the microcalcifications. We used image sizes of 800x800, however, other strategies to encapsulate smaller details would be necessary. Further image augmentation and methods such as batch normalization and dropout layers may need to be incorporated for better results.

## VII. FUTURE

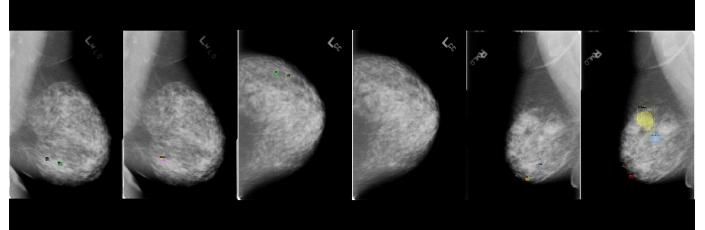
As we mentioned in our proposal, the goal was to compare the results with Momminetv2, as well as use reconstruction losses to enhance graph feature learning. For temporal aspect, we intend to investigate how prior images with different views can be used to enhance the current view for microcalcification images.



(g) Examined, Auxiliary and Contralateral images with ground truth and prediction



(h) Examined, Auxiliary and Contralateral images with ground truth and prediction



(i) Examined, Auxiliary and Contralateral images with ground truth and prediction

## REFERENCES

- [1] Y. Liu, F. Zhang, C. Chen, S. Wang, Y. Wang and Y. Yu, "Act Like a Radiologist: Towards Reliable Multi-view Correspondence Reasoning for Mammogram Mass Detection," in IEEE Transactions on Pattern Analysis and Machine Intelligence, doi: 10.1109/TPAMI.2021.3085783.
- [2] N. Wu, J. Phang, J. Park, Y. Shen, Z. Huang, M. Zorin, S. Jastrzebski, T. Févry, J. Katsnelson, E. Kim et al., "Deep neural networks improve radiologists' performance in breast cancer screening," IEEE transactions on medical imaging, vol. 39, no. 4, pp. 1184–1194, 2019
- [3] Yang Z, Cao Z, Zhang Y, et al. MommiNet-v2: Mammographic multi-view mass identification networks. Med Image Anal. 2021;73:102204. doi:10.1016/j.media.2021.102204
- [4] Abdel-Nasser, Mohamed, Antonio Moreno, and Domenec Puig. "Temporal mammogram image registration using optimized curvilinear coordinates." Computer methods and programs in biomedicine 127 (2016): 1-14.
- [5] Liu, Yuhang, et al. "Cross-view correspondence reasoning based on bipartite graph convolutional network for mammogram mass detection." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020.
- [6] J. Ma, S. Liang, X. Li, H. Li, B. H. Menze, R. Zhang, and W.-S. Zheng, "Cross-view relation networks for mammogram mass detection," arXiv preprint arXiv:1907.00528, 2019
- [7] Y. Li and A. Gupta, "Beyond grids: Learning graph representations for visual recognition," in Advances in Neural Information Processing Systems, 2018, pp. 9225–9235.
- [8] CBIS-DDSM (Curated Breast Imaging Subset of DDSM), <https://wiki.cancerimagingarchive.net/display/Public/CBIS-DDSM>
- [9] INBreast, <https://pubmed.ncbi.nlm.nih.gov/22078258/>
- [10] Breast Micro-Calcification Dataset, <https://zenodo.org/record/5036062>