RESEARCH



Using breast density for hybrid region and pixel-level loss function

Parvaneh Aliniya¹ · Mircea Nicolescu¹ · Monica Nicolescu¹ · George Bebis¹

Received: 15 April 2024 / Revised: 17 March 2025 / Accepted: 21 March 2025 © The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2025

Abstract

Breast cancer continues to be one of the most lethal cancer types, mainly affecting women. However, thanks to the utilization of deep learning approaches, there has been a considerable boost in the performance of the methods for breast cancer detection. The loss function is a core element of any deep learning architecture with a significant influence on their performance. The loss function is particularly important for tasks such as breast mass segmentation. For this task, challenging properties of input images, such as pixel class imbalance, may result in instability of training or poor detection results due to the bias of the loss function toward correctly segmenting the majority class. We propose a hybrid loss function incorporating both pixel-level and region-level losses, where the breast tissue density is used as a sample-level weighting signal. We refer to the proposed loss as Density-based Adaptive Sample-Level Prioritizing (Density-ASP) loss. Our motivation stems from the observation that mass segmentation becomes more challenging as breast density increases. This observation makes density a viable option for controlling the effect of region-level losses. We also propose to evaluate the method using automated density estimation approaches. To demonstrate the effectiveness of the proposed Density-ASP, we conducted mass segmentation experiments using two publicly available datasets: INbreast and CBIS-DDSM. Our experimental results demonstrate that Density-ASP improves segmentation performance compared to the commonly used hybrid losses across multiple metrics.

Keywords Breast cancer \cdot Hybrid loss function \cdot Region-level loss \cdot Pixel-level loss \cdot Density estimation \cdot Pectoral muscle \cdot Mass segmentation \cdot Deep learning

1 Introduction

Despite significant progress in breast cancer screening over the last decade, breast cancer remains one of the most fatal cancer types among women [1]. Mammography is the most common screening tool for breast cancer detection, which has been shown to reduce the mortality rate [2]. Automated

 Parvaneh Aliniya paliniya@unr.edu
 Mircea Nicolescu

mircea@unr.edu

Monica Nicolescu monica@unr.edu

George Bebis bebis@unr.edu breast cancer detection using mammography could help to reduce the cost of a second reader [3, 4] while at the same time increasing the chance of early detection.

Powered by the well-proven effectiveness of deep learning, recent research work on abnormality detection has achieved promising results. However, these methods are still restricted by limitations, such as pixel class imbalance [5], which can adversely affect results in various tasks such as mass segmentation. These limitations often stem from the design of the loss function. In most cases, using a hybrid loss function - the weighted sum of different loss functions - has been shown to be more beneficial compared to nonhybrid loss (using only one type of loss function). Although using the de-facto hybrid loss [5-7] - a weighted sum of Dice [8] and Binary Cross Entropy (BCE) [9] losses - has been shown to provide stability and robustness, the success of recent work on adaptive loss weighting strategies indicates that hybrid loss could be significantly improved by a sample-level design. Following the design of the Adaptive Sample-Level Prioritizing (ASP) loss [10], which uses

¹ Department of Computer Science and Engineering, University of Nevada, Reno, Virginia Street, Reno, NV 89557, USA

the mass ratio in the loss weighting strategy, we propose to employ the breast tissue density associated with each sample in the loss weighting strategy. In this study, we used the breast tissue density defined by the American College of Radiology (ACR) as presented in the datasets. ACR defines the breast density as "a comparison of the relative amounts of fat versus fibroglandular tissue in the breast" [11]. In BIRADS fourth edition [11], there are four density categories: (A) Almost entirely fat, (B) Scattered fibroglandular densities, (C) Heterogeneously dense, and (D) Extremely dense.

The rationale behind selecting breast tissue density as the sample-level signaling feature for hybrid loss stems from the observation that breast density (which represents the composition of fat, fibrous, and glandular tissues) is correlated with mass segmentation's difficulty. In the case of automatic detection from a single view, higher tissue density might be mistaken for an abnormality and increase the false positive rate. Therefore, the breast tissue density of each sample could provide valuable information during training. How to best leverage the information conveyed by each density category for training purposes is an important research question which we attempt to explore in this paper by introducing a region-level loss term in the hybrid loss. The benefit of the region-level loss term lies in the observation that comparison between the regions (rather than pixels) could result in the reduction of false positive and false negative rates by considering the dependencies between the pixels (via including surrounding pixels in the calculation of the loss).

The Density-ASP loss function proposed in this paper consists of pixel-level and region-level losses. In this paper, loss functions such as Dice and BCE that consider pixels independently in the calculation of the loss are referred to as pixel-level losses. On the other hand, the loss functions that take the dependencies between the pixels into consideration are referred to as region-level losses. For the pixel-level loss term, we have combined Dice [5] and BCE [9]. The combination of these losses has been shown to help address the issue of pixel class imbalance and increase training stability [6]. For the region-level loss term, we have combined Structural Similarity Index (SSIM) [12] and Region Mutual Information (RMI) [13]. It should be noted that the term hybrid applies to Density-ASP as well as each of the pixellevel and region-level terms (as they consist of two losses in their own categories). Instead of employing fixed weights [6, 7] for each loss term, following the ASP loss methodology [10], we propose to use the ACR breast density category as an indicative signal for prioritizing the region-level loss term over the pixel-level loss term and vice versa. Therefore, the region-level loss is an adaptive loss term that will be prioritized for samples with higher density.

As mentioned before, breast tissue density is available in the examinations in most publicly available datasets and in the clinical setting. Breast tissue density is useful for two main reasons. First, it is shown that the possibility of development of the cancer is higher when breast tissue density is high [14]. Second, it is more challenging to distinguish the abnormalities from the normal tissues, which leads to unrecognized cancerous tissue and eventually decreases the chance of early detection and successful treatments. As density is assigned by the radiologists, differences of opinion between the radiologists might result in variation between the categories assigned to similar (in terms of density) samples, resulting in inconsistency in the labeling. Therefore, we speculated that this possible inconsistency might affect the results of the proposed method. In order to validate this hypothesis, we utilized a publicly available density estimation approach on the INbreast dataset. It would enable a unified density category labeling for all the samples. The Openbreast [15] software was used for automated density estimation. The presence of the pectoral muscle, which has a high brightness similar to breast tissues, makes the estimation more challenging; hence, to mitigate this effect, the pectoral muscle is removed before density estimation. In addition, the baseline method was also tested using images in which pectoral muscle was removed. Using AU-Net [6], which is a modern and effective variation of U-Net [16], as the baseline architecture, the Density-ASP loss has been evaluated on two benchmark datasets for mass segmentation: INbreast [17], and CBIS-DDSM [18]. The results of our experiments illustrate that Density-ASP loss provides considerable performance improvements compared to commonly used hybrid losses.

The contributions of this paper are listed below:

- Incorporating both pixel-level and region-level losses in the Density-ASP loss function.
- Employing breast tissue density as a prioritizing signal for adaptive sample-level prioritizing loss function for mass segmentation on whole mammograms.
- Evaluating Density-ASP on two benchmark datasets, INbreast and CBIS-DDSM.
- Performing ablation studies on the INbreast dataset to measure the effect of using automatically generated density labels as a prioritizing signal.
- Measuring the effect of removing pectoral muscle in the experiments.
- Quantitatively analyzing and comparing the findings of our experimental results for Density-ASP loss with the traditional hybrid loss for the baseline approach and state-of-the-art mass segmentation methods.

In the following sections, we first review related work in the field. Then, we explain the proposed method in detail. Next, we present our experimental results, analysis, ablation study, and comparison with state-of-the-art methods. Finally, we provide our conclusions and future research directions.

2 Related work

Recently, deep learning-based approaches have shown great promise in abnormality detection in medical images, with many studies achieving more accurate mass segmentation results compared to traditional approaches. In this section, we aim to briefly review the related work in deep learningbased approaches for breast mass segmentation, categorizing them into two groups: breast mass segmentation in whole mammograms and loss functions for binary segmentation of medical images.

2.1 Mass segmentation on whole mammograms

The majority of breast mass segmentation approaches fall into one of the following categories based on the type of input they receive: region of interest (RoI) and whole mammogram. RoI-based mass segmentation approaches [19] have different properties, challenges, and strategies compared to methods using whole mammograms [6, 7]; thus, in this section, the primary focus is on reviewing related work in the latter category.

Inspired by [20], one of the pioneer deep learning-based approaches for segmentation, Ronneberger et al. proposed U-Net [16], which is a fully convolutional symmetric encoder-decoder architecture that is instrumental for segmentation tasks with limited training data. This property of U-Net makes it specifically favorable for medical image segmentation, where data scarcity is a relatively common limitation. U-Net combines low-level location information from the encoder with high-level semantic information from the decoder.

Thanks to the effectiveness of U-Net, a new wave of variations for different medical tasks has emerged [21–27], continuing to push the performance boundaries of medical image segmentation. In this context, the method proposed in [28] introduced a similar encoder-decoder architecture (leveraging dense blocks) where multi-scale information is utilized in the network. To enhance the performance of the network without additional parameters, atrous convolution [29] with various sample rates was used in the last encoder block. Li et al. [30] proposed another U-Net-based approach based on the idea of utilizing a densely connected network in the encoder and a CNN with attention gates in the

decoder. Another line of research within the scope of multiscale studies is [31], where the generator was designed as an improved version of U-Net. Before sending the segmentation results to the discriminator, multiscale results were created for three critics with different scales in the discriminator. Ravitha et al. [32], developed an approach employing the error of the outputs of intermediate layers relative to the ground truth labels as a supervision signal to boost the model's performance.

Sun et al. [6] introduced an attention-guided dense-upsampling asymmetric encoder-decoder network (AU-Net) with an intermediate up-sampling block, which includes a channel-wise attention mechanism designed to leverage the beneficial information presented in both low and high-level features. To mitigate the problem of relatively low performance of the U-Net approach on small-size masses, Xu et al. [7] proposed to use a selective receptive field module with two parts, one for generating several receptive fields with different sizes and one for selecting the appropriate size of the receptive field. AU-Net has been chosen as the baseline model in this study.

2.2 Loss for medical image segmentation

The choice of a suitable loss function, conveying the desired objectives of the task performed by a network, has a tremendous impact on the training process and overall performance of the network. Among the previously introduced losses for segmentation, while some consider the pixels independent entities, others seek to take regional information into consideration to capture the dependencies between the pixels. The first group is generally regarded as pixel-level losses, and the latter as region-level losses in the literature. Considering that both categories are relevant to this research, we provide a concise summary of related studies in both groups, starting with pixel-level approaches and emphasizing the ones proposed for the medical domain.

Binary segmentation could be considered as the classification of pixels into positive (foreground) and negative (background) classes. A common loss function for this task is the BCE loss [9] (Eq. 1), which penalizes the discrepancy between predicted and ground truth classes for all pixels. Weighted Binary Cross Entropy [33] and Balanced Cross Entropy [34] are two BCE variants that differentiate between the effect of false positives and false negatives through weighting coefficients. Focal loss [35] further improved BCE by changing the magnitude of the loss according to the hardness of the examples based on the confidence of the model. Dice loss [8] is suitable for addressing the pixel class imbalance problem [36], formulated as the ratio of correctly classified pixels to the total number of positive pixels in the prediction and ground truth masks (Eq. 2). Tversky loss [37] provides a way to control the contribution of the false positive and the false negative terms in the Dice loss by weighting these terms.

All the aforementioned losses belong to the pixel-level category (i.e., they consider the pixels independently). While providing effective training signals for the network, they neglect to consider the relationship among pixels, which could provide a considerable boost, notably for cases with irregularity in shapes. Initially proposed for image quality assessment, SSIM [12] has been incorporated in the segmentation loss for medical image segmentation [38] and has inspired several region-level losses. SSL [39] and RMI [13] are two examples of region-level losses developed for segmentation. It should be noted that both of these losses consider a fixed-size window around each pixel as the region (a region is defined for each pixel) rather than a fixed location (a region is a fixed location in a grid) in the ground truth and the prediction as utilized in [40, 41].

SSIM [12] uses luminance, contrast, and structure in measuring the differences between two regions. Inspired by the influence of the structural term in the SSIM, which has the potential to be customized for segmentation purposes, the authors of Structural Similarity Loss (SSL) [39] proposed to weight the cross-entropy of every two pixels based on the structural error (error between two image regions which indicates the degree of linear correlation) while ignoring pixels with low error and emphasizing on pixels with high error by thresholding the error rate. With the goal of maximizing the structural similarity between images, RMI [13] first converts the region around a center (pixel) to a multi-dimensional point (for a 3×3 region, it will be a 9D point) and then maximizes the MI between multidimensional distributions.

Several compound losses [35, 37, 42] have been proposed to reap the benefits of different losses by combining two or more of them. Combo loss [42] has been proposed to control the contribution of false positive and false negative by a weighting strategy in the BCE loss term where the total loss is a weighted sum of BCE and Dice loss. In adaptive samplelevel prioritizing loss, we have proposed a novel approach to weight the loss terms (Dice and BCE) dynamically. This is performed in an adaptive manner by controlling the influence of each loss according to each sample using the ratio of the mass to image size as a weighting signal. It should be noted that ASP has three versions: quantile-based, clusterbased, and learning-based. In the quantile-based ASP, the images are grouped based on the quantile to which the ratio of the mass belongs. In the cluster-based version, the category of an image is identified according to the K-means clustering of the ratios. Finally, the learning-based ASP is a parametrized version of the ASP loss.

We follow the same dynamic weighting strategy as ASP loss [10]. However, instead of utilizing the size of the mass, we opt to utilize the breast tissue density, which is mostly related to the difficulty of the segmentation. In addition, instead of solely using and weighting pixel-level losses, we introduce a combination of hybrid pixel-level and region-level losses.

2.3 Pectoral muscle removal

Pectoral muscle removal was utilized as a preprocessing step for the automated density estimation and in ablation studies for the proposed method. Hence, in this section, we provide a review of the previous methods proposed to tackle the pectoral muscle removal task. According to a recent study [43], thresholding [44–47] and region growing [48–50] are among widely used approaches for pectoral muscle removal.

The core idea for thresholding is based on the observation that the brightness of the pectoral muscle is generally higher than the neighboring regions; therefore, by eliminating pixels lower than a certain threshold, the region for the pectoral muscle will the extracted. This idea, coupled with the utilization of the orientation of the breast (whether the breast is on the left or right side of the image) and the generic shape of the muscle, has been used in the literature. In this category, Subashini et al. [45] used a thresholdingbased approach for pectoral muscle removal in which they first extracted the rectangle in the image where the pectoral muscle was assumed to be located and then used thresholding within the rectangle to detect the muscle. Tayel et al. [46] proposed an approach that eliminates the need for a predefined region. In the same category, Czaplicka et al. [47] proposed to use a multi-level thresholding method, and Shrivastava et al. [51] developed a method using sliding window for thresholding.

The second category of methods for pectoral muscle removal consists of region-growing-based methods. In general, these methods start with initial seeds; then, according to certain similarity metrics, they continue adding a new neighboring pixel to a region until a termination criterion is met [48]. Chen et al. [52] proposed to use a pixel near the border between the pectoral muscle and the breast tissue (which was approximated) as a starting seed. Instead of using the approximation of the location for the border, Nagi et al. [53] used the approximated location for the pectoral muscle after determining the orientation for the breast to place the initial seed. Maitra et al. [54] introduced several improvements to the previous method. For seed selection, they proposed to use the diagonal of a defined rectangle encapsulating the pectoral muscle from top-left to rightbottom. In addition, they used new selection criteria based

on the minimum, maximum, and average values for the pixels. Aside from the previous methods that aim to use region growing and thresholding, graph-cut [55], Hough Transform [56], line estimation, polynomial fitting, and curve estimation [57], k-means [58], active contours [59] and contour growing [60] are also used in several methods.

2.4 Automated breast tissue density estimation

According to [61], breast density could be considered as an independent risk factor or in addition to other measurements [62] for the development of breast cancer. In this section, an overview of recent automated breast tissue density estimation approaches has been provided. Breast density could be defined as the percentage of the fibroglandular tissue relative to fat. In the literature, the "parenchymal pattern" and "percent density" are also used to refer to the same concept [61]. In the BI- RADS version 5, the definition is changed in a way that percentage is no longer applied in the form of numerical values [11] due to the mismatch between the automated volume calculation and assessment of radiologists. The latest version relies only on the visual assessment of the radiologists [11]. This, in turn, could introduce inconsistency between the density category for automated approaches in which the assumption is that the definition of the density category is universal for all the samples, while in reality, visual evaluation of several radiologists might introduce differences in the categories for samples with similar visual appearance; therefore, the performance of the methods might suffer. In this regard, the proposed approaches could be divided into two groups: supervised and unsupervised approaches.

The supervised approaches aim to automate the density estimation and use the radiologist's assessment of the density to train the models. In this category, Lehman et al. [63] proposed the use of deep learning for the classification of the samples. Saffari et al. [64] proposed to use conditional-GAN [65] for training a model to generate segmentation of the dense tissue, then the output of the cGAN has been classified using the classification network.

The methods in the unsupervised category rely only on the input images for density estimation, mostly by utilizing the handcrafted features such as thresholding [66], region growing [67] and clustering [68]. Kallenberg et al. [69] proposed a CNN network to use unlabeled data for the classification of images. Qyantra [70], Openbreast [15], Libra [68], and Volpara [71] are among the software that have a breast density estimation system. In the ablation study in this paper, we have used Openbreast software as the automated breast density classifier.

3 Methodology

In this study, both hybrid pixel-level and region-level loss functions are utilized. Therefore, an overview of hybrid pixel-level L_{HP} and region-level L_{HR} losses are provided in the following sections.

3.1 Hybrid pixel-level loss

The standard hybrid pixel-level loss, commonly used for mass segmentation in mammograms, is defined as a weighted sum of BCE and Dice loss, as shown below:

$$L_{BCE} = -\left(y log(\hat{y}) + (1 - y) log(1 - \hat{y})\right),$$
 (1)

$$L_{Dice} = 1 - \frac{\sum_{j=1}^{H \times W} \hat{y}_j y_j + \epsilon}{\sum_{j=1}^{H \times W} \hat{y}_j + \sum_{j=1}^{H \times W} y_j + \epsilon},$$
 (2)

$$L_{HP} = \alpha L_{Dice} + \beta L_{BCE}.$$
(3)

Here y and \hat{y} represent the ground truth and the predicted segmentation masks, respectively. α and β (could be relative, for instance, formulated as $\beta = 1 - \alpha$) are the weighting parameters in the hybrid loss denoted as L_{HP} in Eq. 3. While the cross entropy loss (Eq. 1) includes correctly classified positive and negative pixels, the Dice loss (Eq. 2) incorporates only correctly classified positive pixels, which makes it more suitable in the presence of considerable pixel class imbalance. The combination (Eq. 3) of the two losses has been shown to provide a better learning signal. In particular, it has been reported that adding BCE to the Dice loss helps to mitigate the unstable training associated with using only the Dice loss [35, 36]. On the other hand, adding the Dice loss to BCE helps to improve the performance of the model on datasets with pixel class imbalance compared to using BCE alone.

3.2 Hybrid region-level loss

Region-level losses aim to incorporate the context to which a pixel belongs in the loss calculation by representing each pixel with its own value and the neighboring pixels' values. In this paper, two of the region-level losses, SSIM and RMI, have been selected and are represented in the following:

$$L_{RMI}(Y_m; \hat{Y_m}) = \int_S \int_{\hat{S}} f(y, \hat{y}) \log\left(\frac{f(y, \hat{y})}{f(y)f(\hat{y})}\right) \, dy \, d\hat{y}.$$
 (4)

Here, Y_m and $\hat{Y_m}$ are the multi-dimensional points constructed using a centering pixel and the neighboring pixels in a surrounding square. S and \hat{S} are the support sets corresponding to the ground truth and prediction masks, respectively. f(y) and $f(\hat{y})$ represent the probability density functions for the ground truth and prediction masks in Eq. 4, respectively. The $f(y, \hat{y})$ captures the joint PDF. The implementation details of the RMI loss are available in [13]. The second region-level loss used in this paper is SSIMbased loss in Eq. 5.

$$L_{SSIM}(Y_p; \hat{Y}_p) = 1 - \frac{(2\mu_{y_p}\mu_{\hat{y}_p} + C_1)(2\sigma_{y\hat{y}} + C_2)}{(\mu_{y_p}^2 \mu_{\hat{y}_p}^2 + C_1)(\sigma_y^2 + \sigma_{\hat{y}}^2 + C_2)}.$$
(5)

Here, Y_p and \hat{Y}_p represent patches in the ground truth and the prediction masks. μ and σ are the mean and variance for the corresponding patches, respectively. $\sigma_{y\hat{y}}$ is the covariance of the two patches. More details (including the selection of C1 and C_2) are available in [12]. Finally, the hybrid region-level loss is presented in Eq. 6.

$$L_{HR} = \eta L_{RMI} + \gamma L_{SSIM}.$$
 (6)

In the hybrid region-level loss L_{HR} (Eq. 6), L_{RMI} and L_{SSIM} are the RMI and SSIM losses, respectively. The hyperparameters η and γ represent the weighting coefficients.

3.3 Density-adaptive sample-level prioritizing loss

While the aforementioned hybrid pixel-level loss is quite effective, we propose extending it by using an adaptive weighting strategy based on the idea of ASP [10]. The resulting hybrid loss is a combination of region-level and pixel-level losses instead of using only pixel-level losses. We propose using breast tissue density as the sample-level signal for the extended hybrid loss's prioritizing strategy. In the following, the proposed framework for the Density-ASP loss is explained. Given a training set of N images and

the corresponding segmentation masks, the baseline method learns a mapping between an input image to its segmented counterpart using the training data. In this study, AU-Net was selected as the baseline method; the architecture for AU-Net is presented in Fig. 1. For the encoder and decoder, ResUnit and the basic decoder proposed in AU-Net have been used. The details of the AU block, basic decoder, and ResUnit encoder are presented in [6]. The Density-ASP loss requires the breast tissue density for each sample. The standard ACR density, which is available in both datasets, was used in this study. ACR breast density reflects the composition of the fat, fibrous, and glandular tissue in four categories.

There are noticeable differences in the appearance of the breast within different density categories in mammography images. Generally, the complexity of the texture increases as density increases. This provides meaningful distinguishing information for the loss function to prioritize the pixellevel or region-level terms in the loss function based on the density of each sample. The more complex the texture is (i.e., higher density category), the more important the contribution of the region-level term will be. Therefore, density is considered a determining factor in the weighting strategy.

In Fig. 1, the prediction heatmap (\hat{y} in the formulas) and ground truth segmentation masks are inputted to the Density-ASP module. The process of prioritizing loss is presented in Fig. 2. Since there is no proven or intuitive connection between density and pixel-level terms, the weight for this category remains fixed. However, the contribution of the region-level term will change in an adaptive manner, as shown in Eq. 7 and Fig. 2. It should be noted that the weighting coefficients inside the pixel-level and region-level loss terms are not adaptive.

$$L^{i}_{Density-ASP} = p^{i} \theta^{T} L^{i}_{HR} + L^{i}_{HP}.$$
(7)



Fig. 1 The overall architecture of the proposed method



Fig. 2 Density-ASP Block

 $L_{Density-ASP}^{i}$ is the final Density-ASP loss for the *i*th sample. θ is the prioritizing vector consisting of the weights assigned to each density category, and p_i denotes a one-hot encoding of the density category to which the *i*th sample belongs. $p^i \theta^T$ will be the weight for the region-level loss term, which determines the importance of the region-level term according to the density category of *i*th sample.

4 Experimental results

This section begins with a description of the datasets and evaluation metrics. Subsequently, the experimental setting is presented, followed by a comprehensive analysis of the results on both datasets, including comparisons with the state-of-the-art approaches.

4.1 Datasets

We have conducted mass segmentation experiments using two publicly available datasets: INbreast and CBIS-DDSM. We have normalized the intensity of the images in both datasets, and all images have been resized to 256×256 . No data augmentation or image enhancement was considered in our experiments. To prevent overfitting, a randomly selected validation set was utilized for hyperparameter tuning. For the baseline approach, the batch size was set to four, the learning rate was initially set to 10–e4, and a step decay policy with a decay factor of 0.5 was employed in all experiments. Irrespective of the abnormality type, all the images containing masses have been utilized in our experiments.

4.1.1 INbreast dataset

The INbreast dataset contains 410 images associated with 150 cases, including various abnormality types. In the context of mass segmentation, only 107 of the images containing masses (the total number of masses across all of the images is 116) have been used in this study. A 5-fold cross-validation was employed, a commonly used setting for the measurement of the performance of methods on the INbreast dataset. The dataset was randomly divided into training (80%), validation (10%), and test (10%) sets.

4.1.2 CBIS-DDSM dataset

From a total of 1944 cases in the CBIS-DDSM dataset, 1591 images containing masses were utilized in our experiments. The official split of the dataset (1231 and 360 images for train and test sets, respectively) was employed for the experimental results presented in this paper. 10% of the training data was randomly selected for the validation set. In a preprocessing stage for the CBIS-DDSM dataset, artifacts were removed, and images were cropped and resized.

4.2 Evaluation metrics

Since mass segmentation in mammograms is characterized by a pixel class imbalance, we have selected several metrics to better illustrate the strengths and weaknesses of the proposed methods. Specifically, the Dice Similarity Coefficient (DSC), Relative Area Difference (ΔA), Sensitivity, and Accuracy have been selected due to the complementary information they provide. This combination of evaluation metrics highlights the performance of each method both in majority (background) and minority (masses) classes. It also reflects how accurately a method performs in terms of predicting the boundary of masses, which is crucial for mass classification.

4.3 Ablation study

Many design factors impact the performance of a deep learning-based method. In this section, we explore several design factors specifically related to the proposed method. As the method's goal is to enhance the segmentation results for images with higher density, we design several tests to select the preprocessing steps accordingly.

Firstly, we tested using U-Net and Au-Net as the baseline model; since the results for the AU-Net seemed to be more promising, we opted to perform all the experiments using the AU-Net, specifically because it was initially proposed to improve the segmentation performance on mammography **Fig. 3** Examples of original images (first row) and images in which the pectoral muscle was removed (second row)



Table 1 Results for experiments using pectoral muscle removal for INbreast

	DSC	ΔA	SEN	ACCU
Whole image	65.32	23.68	57.95	98.46
After pectoral muscle removal	45.67	2.06	51.48	95.59

Table 2 Results for experiments using automated density estimation for INbreast

	DSC	ΔA	SEN	ACCU
Ground-truth density	65.32	23.68	57.95	98.46
Density from openbreast	70.74	12.25	71.76	98.48
I				

images. In general, the baseline method could be a different method as well.

The second experiment involves using original images or preprocessing them with pectoral muscle removal that affects the MLO images. Finally, the third experiment targets the impact of using automated density estimation instead of relying on the labels provided by different radiologists.

4.3.1 Pectoral muscle removal in the preprocessing

The pectoral muscle, present in the MLO view images, is a portion of the image that is not part of the breast but has brightness similar to more dense regions in the breast. This, in turn, could introduce errors in the performance of the methods. In order to investigate the effect of the presence of pectoral muscle, we considered using pectoral muscle removal in the preprocessing step to determine the effect and necessity of the operation in the final results. Ground truth labels provided by Aliniya et al. [72] were utilized as the masks for removing pectoral muscle. As the labels provided in [72] are manually determined, they have high accuracy (Fig. 3). Table 1 presents the results for the training of the baseline model using the dataset in which the pectoral muscle was removed and the original images. As shown in the table, generally the performance is lower compared to the baseline method. For this reason, we did not use pectoral

muscle removal in the preprocessing step of the proposed method.

4.3.2 Using automated density estimation approach to unify the density labels

As previously mentioned, the differences between assessments of several radiologists might negatively impact the performance of the proposed method. To investigate these effects, we propose to use an automated density estimation approach to achieve a unified density category for all the images in the dataset. The results for these experiments have been presented in Table 2. The percent density is converted to four categories after thresholding. As shown, the performance of the proposed method has improved compared to the baseline method using the density labels generated by Openbreast for the INbreast dataset; however, it is less than that of ground-truth labels. It indicates that the ground-truth labels are more useful for the proposed method. The reason might be that as the performance of these approaches is limited, there is a possibility of the propagation of the error from the density estimation method to the mass segmentation approach, which affects the results adversely.

4.4 Comparison with state-of-the-art methods

To assess the performance of Density-ASP loss, we have conducted a comprehensive comparison with three state-ofthe-art mass segmentation approaches on whole mammograms: AU-Net (baseline), ARF-Net, and ASP. The official implementation of AU-Net, and the setting described in the AU-Net paper [6] (only the architecture was publicly available) were used in our experiments. ARF-Net is a state-of-the-art method for mass segmentation on whole mammograms. The method was implemented to the best of our understanding based on the original paper (i.e., the implementation of the approach or the trained models were not publicly available). For the ASP loss, we have used the same experimental setting and data split, so we have directly used the reported results in the original paper. The publicly available implementations of the RMI and SSIM were utilized. To ensure a fair comparison of the methods, no pre-training or data augmentation was used. The coefficients for density-based loss were $\theta = [0.5, 0.5, 0.85, 0.95]$ and $\theta = [0.25, 0.25, 0.85, 0.95]$ for the INbreast and CBIS-DDSM, respectively. η , γ , β were set to one; α was set to 2 and 2.5 for INbreast and CBIS-DDSM. These hyperparameters were selected through experimental evaluation.

4.4.1 Experimental results using INbreast

Table 3 summarizes our experimental results for all models trained on INbreast. The best results are highlighted using bold font. The Density-ASP loss achieved better performance across all of the metrics compared to the pixellevel hybrid losses. The improvement for the Density-ASP (over using hybrid pixel-level loss in the baseline method) is as follows: (DSC: +9.27%, ΔA : -12.77%, Sensitivity: +20.21%, Accuracy: +0.19%), which is consistent across all metrics. The Density-ASP outperformed ARF-Net in DSC, ΔA , and sensitivity while the accuracy is 0.06% lower. It should be noted that ARF-Net is designed to incorporate different sizes and, surpasses the baseline method in DSC, sensitivity, and accuracy. Better performance of the Density-ASP (in most metrics) compared to ARF-Net, indicates that improvement in the training that Density-ASP provides for the baseline method not only closes the gap between

 Table 3 Results for Density-ASP and state-of-the-art approaches for INbreast

Method	$\mathrm{DSC}\uparrow$	$\Delta A\downarrow$	Sensitivity ↑	Accuracy ↑
ARF-Net	70.05	30.37	59.59	98.71
AU-Net (baseline)	65.32	23.68	57.95	98.46
ASP-Quantile-based	68.03	25.04	63.12	98.54
ASP-Learning-based	71.92	22.31	64.56	98.71
ASP-Cluster-based	74.18	19.28	67.21	98.78
Density-ASP	74.59	10.91	78.16	98.65

AU-Net and ARF-Net in most of the metrics but also makes AU-Net outperform ARF-Net. In comparison with the ASP loss variations (as the best-performing version, cluster-based ASP was selected for comparison), Density-ASP performed better in terms of DSC, ΔA , and sensitivity. The accuracy of the cluster-based ASP variation is 0.13% higher than the Density-ASP. The results of the Density-ASP further validate the effectiveness of sample-level losses. Moreover, the fact that Density-ASP outperforms the ASP in most of the metrics indicates that introducing the region-level losses to the loss function with the density as a weighting signal is a promising approach for mass segmentation. We attribute this improvement to the utilization of density as the prioritizing signal, which helps to distinguish the contribution of the losses for each sample, leading to better segmentation.

The first four columns in Fig. 4 show some representative results for Density-ASP, AU-Net, ARF-Net, and all ASP variations for INbreast. These examples have been selected to include instances for each density category (mentioned at the top of the columns), demonstrating the segmentation capabilities of the methods across different density categories. The green and blue lines represent the contours of the ground truth and the prediction masks, respectively. It can be observed that the segmentation results for Density-ASP are more accurate compared to state-of-the-art methods across all the density categories.

4.4.2 Experimental results using CBIS-DDSM

The performance of Density-ASP loss on the CBIS-DDSM dataset compared to state-of-the-art approaches is presented in Table 4. The improvement for the Density-ASP on the CBIS-DDSM dataset (over using hybrid pixel-level loss in the baseline method) is as follows: (DSC: +1.59%, ΔA : -3.98%, Sensitivity: +0.66%, Accuracy: +0.03%). The Density-ASP outperformed ARF-Net (which has a different architecture but used common hybrid loss) in all metrics except for accuracy, which is 0.02% lower. When compared to quantile-based ASP loss - a version of the ASP with the best performance on CBIS-DDSM - while Density-ASP outperformed quantile-based ASP in sensitivity (+0.15%)), it under-performed in other metrics (DSC: -0.4%, ΔA : +3.91%, Accuracy: -0.03%). We speculate that the reason might be related to the fact that the mass ratio in ASP loss is a data-driven factor that completely correlates with the statistics of the pixels in the image. On the other hand, density is predefined and, in some cases, might not be aligned with the visual features, which might be a more common issue in the CBIS-DDSM dataset. The fact that Density-ASP improves in all of the metrics over the baseline method shows the effectiveness of using density and region-level losses for the CBIS-DDSM dataset. The last four columns in



Fig. 4 Examples of the segmentation results for Density-ASP, AU-Net, ARF-Net, and ASP variations

CBIS-E	DDSM						
Table 4	Results	for	Density-ASP	and	state-of-the-art	approaches	for

Method	$\mathrm{DSC}\uparrow$	$\Delta A\downarrow$	Sensitivity \uparrow	Accuracy ↑
ARF-Net	48.82	11.47	47.27	99.43
AU-Net (baseline)	49.05	09.94	51.49	99.38
ASP-Quantile-based	51.48	02.05	52.00	99.43
ASP-Learning-based	51.33	23.17	45.38	99.50
ASP-Cluster-based	51.04	04.47	49.90	99.45
Density-ASP	50.64	05.96	52.15	99.41

Bold font is used to highlight the best performance among methods for each metric

Fig. 4 show some representative examples where Density-ASP has better performance when compared to the previous methods in different density categories.

In general, the performance of the Density-ASP is better for the INbreast dataset. One observation is that in both datasets, there are examples that the density category of the image might not be perfectly aligned with visual features (for example, the 2nd column in Fig. 4), which could cause a higher weight for the term that does not match the initial idea of the Density-ASP. We speculate that the assigned density for the INbreast is more visually aligned with the images, resulting in better weighting for loss terms. Different distributions of each category in the datasets might also be a factor in the performance of the Density-ASP on two datasets.

5 Conclusion

We have proposed a new sample-level adaptive prioritizing loss that utilizes breast tissue density as a weighting signal. Moreover, we have proposed a hybrid loss function that includes region-level losses in the training. Finally,

given the observed connection between the difficulty of mass identification and the breast tissue density category, this approach focuses on using the density for weighting of the region-level loss term to highlight the importance of the region-level term according to the density category for each sample adaptively. Our experimental results demonstrate improvements in all evaluation metrics on two benchmark datasets: INbreast and CBIS-DDSM. Customizing this category of losses for other domains or tasks is an appealing direction for future work. In the ablation study, we explored the idea of using automated density labels and removing the pectoral muscle in the preprocessing steps. The pectoral muscle removal did not provide improvement over the baseline which could be due to the fact that it increases the background pixels. While outperforming the baseline method, the automated density estimation did not provide an improvement over using the ground truth labels. The reason might be that the method relies on thresholding for the classification of the density. Using automated density estimation methods that are independent of thresholds might improve the performance.

Declarations

Conflict of interest The authors declare that have no conflict of interest.

References

- Siegel, R.L., Miller, K.D., Fuchs, H.E., Jemal, A.: Cancer statistics, 2022. CA: A Canc. J. Clin. 72(1), 7–33 (2022)
- Nyström, L., Andersson, I., Bjurstam, N., Frisell, J., Nordenskjöld, B., Rutqvist, L.E.: Long-term effects of mammography screening: updated overview of the Swedish randomised trials. The Lancet **359**(9310), 909–919 (2002)
- 3. Batchu, S., Liu, F., Amireh, A., Waller, J., Umair, M.: A review of applications of machine learning in mammography and future challenges. Oncology **99**(8), 483–490 (2021)
- McKinney, S.M., Sieniek, M., Godbole, V., Godwin, J., Antropova, N., Ashrafian, H., Back, T., Chesus, M., Corrado, G.S., Darzi, A., Etemadi, M.: International evaluation of an AI system for breast cancer screening. Nature 577(7788), 89–94 (2020)
- Malof, J.M., Mazurowski, M.A., Tourassi, G.D.: The effect of class imbalance on case selection for case-based classifiers: An empirical study in the context of medical decision support. Neural Netw. 25, 141–145 (2012)
- Sun, H., Li, C., Liu, B., Liu, Z., Wang, M., Zheng, H., Feng, D.D., Wang, S.: AUNet: attention-guided dense-upsampling networks for breast mass segmentation in whole mammograms. Phys. Med. Biol. 65(5), 055005 (2020)
- Xu, C., Qi, Y., Wang, Y., Lou, M., Pi, J., Ma, Y.: ARF-Net: An Adaptive Receptive Field Network for breast mass segmentation in whole mammograms and ultrasound images. Biomed. Signal Process. Control 71, 103178 (2022)
- 8. Milletari, F., Navab, N., Ahmadi, S.-A.: V-net: Fully convolutional neural networks for volumetric medical image segmentation. In:

2016 fourth international conference on 3D vision (3DV), IEEE, pp. 565–571 (2016)

- Yi-de, M., Qing, L., Zhi-Bai, Q.: Automated image segmentation using improved PCNN model based on cross-entropy. In: Proceedings of 2004 International Symposium on Intelligent Multimedia, Video and Speech Processing, IEEE, pp. 743-746 (2004)
- Aliniya, P., Nicolescu, M., Nicolescu, M., Bebis, G.: ASP Loss: Adaptive Sample-Level Prioritizing Loss for Mass Segmentation on Whole Mammography Images. In: International Conference on Artificial Neural Networks (ICANN), (2023)
- Spak, D.A., Plaxco, J.S., Santiago, L., Dryden, M.J., Dogan, B.E.: BI-RADS®fifth edition: A summary of changes. Diagn. Interv. Imaging 98(3), 179–190 (2017)
- Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. 13(4), 600–612 (2004)
- Zhao, S., Wang, Y., Yang, Z., Cai, D.: Region mutual information loss for semantic segmentation. Advances in Neural Information Processing Systems 32 (2019)
- Nara, M., Fujioka, T., Mori, M., Aruga, T., Tateishi, U.: Prediction of breast cancer risk by automated volumetric breast density measurement. Jpn. J. Radiol. 41(1), 54–62 (2023)
- Pertuz, S., Torres, G.F., Tamimi, R., Kamarainen, J.: Open framework for mammography-based breast cancer risk assessment. In: 2019 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI). IEEE, pp. 1–4 (2019)
- Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015, Springer International Publishing, pp. 234–241 (2015)
- Moreira, I.C., Amaral, I., Domingues, I., Cardoso, A., Cardoso, M.J., Cardoso, J.S.: Inbreast: toward a full-field digital mammographic database. Acad. Radiol. 19(2), 236–248 (2012)
- 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. Sci. Data 4(1), 1–9 (2017)
- Baccouche, A., Garcia-Zapirain, B., Castillo Olea, C., Elmaghraby, A.S.: Connected-UNets: a deep learning architecture for breast mass segmentation. NPJ Breast Canc. 7(1), 151 (2021)
- Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 3431–3440 (2015)
- Wu, S., Wang, Z., Liu, C., Zhu, C., Wu, S., Xiao, K.: Automatical segmentation of pelvic organs after hysterectomy by using dilated convolution u-net++. In: 2019 IEEE 19th international conference on software quality, reliability and security companion (QRS-C). IEEE, pp. 362–367 (2019)
- Zhang, J., Jin, Y., Xu, J., Xu, X., Zhang, Y.: Mdu-net: Multi-scale densely connected u-net for biomedical image segmentation. arXiv preprint arXiv:1812.00352 (2018)
- Zhou, Z., Rahman S., Md Mahfuzur, Tajbakhsh, N., Liang, J.: Unet++: A nested u-net architecture for medical image segmentation. In: Deep learning in medical image analysis and multimodal learning for clinical decision support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4, Springer International Publishing, pp. 3–11 (2018)
- 24. Li, C., Tan, Y., Chen, W., Luo, X., Gao, Y., Jia, X., Wang, Z.: Attention unet++: A nested attention-aware u-net for liver CT image segmentation. In: 2020 IEEE international conference on image processing (ICIP), IEEE, pp. 345–349 (2020)

- Cao, H., Wang, Y., Chen, J., Jiang, D., Zhang, X., Tian, Q., Wang, M.: Swin-unet: Unet-like pure transformer for medical image segmentation. In: European conference on computer vision, Cham: Springer Nature Switzerland, pp. 205-218 (2022)
- Oktay, O., Schlemper, J., Folgoc, L.L., Lee, M., Heinrich, M., Misawa, K., Mori, K., et al. Attention u-net: Learning where to look for the pancreas. arXiv preprint arXiv:1804.03999 (2018)
- Song, T., Meng, F., Rodriguez-Paton, A., Li, P., Zheng, P., Wang, X.: U-next: A novel convolution neural network with an aggregation u-net architecture for gallstone segmentation in CT images. IEEE Access 7, 166823–166832 (2019)
- Hai, J., Qiao, K., Chen, J., Tan, H., Xu, J., Zeng, L., Shi, D., Yan, B.: Fully convolutional densenet with multiscale context for automated breast tumor segmentation. J Healthc 2019(1), 8415485 (2019)
- Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L.: Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE Trans. Pattern Anal. Mach. Intell. 40(4), 834–848 (2017)
- Li, S., Dong, M., Du, G., Mu, X.: Attention dense-u-net for automatic breast mass segmentation in digital mammogram. IEEE Access 7, 59037–59047 (2019)
- Chen, J., Chen, L., Wang, S., Chen, P.: A novel multi-scale adversarial networks for precise segmentation of x-ray breast mass. IEEE Access 8, 103772–103781 (2020)
- Rajalakshmi, N.R., Vidhyapriya, R., Elango, N., Ramesh, N.: Deeply supervised u-net for mass segmentation in digital mammograms. Int. J. Imaging Syst. Technol. 31(1), 59–71 (2021)
- Pihur, V., Datta, S., Datta, S.: Weighted rank aggregation of cluster validation measures: a monte carlo cross-entropy approach. Bioinformatics 23(13), 1607–1615 (2007)
- Xie, S., Tu, Z.: Holistically-nested edge detection. In: Proceedings of the IEEE international conference on computer vision, pp. 1395–1403 (2015)
- Yeung, M., Sala, E., Schönlieb, Carola-Bibiane., Rundo, L.: Unified focal loss: generalising dice and cross entropy-based losses to handle class imbalanced medical image segmentation. Computer. Med. Imag. Graph. 95, 102026 (2022)
- Jadon, S.: A survey of loss functions for semantic segmentation. In: 2020 IEEE conference on computational intelligence in bioinformatics and computational biology (CIBCB), pp. 1-7. IEEE (2020)
- Salehi, S.S.M., Erdogmus, D., Gholipour, A.: Tversky loss function for image segmentation using 3D fully convolutional deep networks. In: Machine Learning in Medical Imaging: 8th International Workshop, MLMI 2017, Proceedings 8. Springer International Publishing, pp. 379–387 (2017)
- Huang, H., Lin, L., Tong, R., Hu, H., Zhang, Q., Iwamoto, Y., Han, X., Chen, Y.-W., Wu, J.: Unet 3+: A full-scale connected unet for medical image segmentation. In: ICASSP 2020-2020 IEEE international conference on acoustics, speech and signal processing (ICASSP). IEEE, pp. 1055–1059 (2020)
- Zhao, S., Wu, B., Chu, W., Hu, Y., Cai, D.: Correlation maximized structural similarity loss for semantic segmentation. arXiv preprint arXiv:1910.08711 (2019)
- Aliniya, P., Razzaghi, P.: Parametric and nonparametric context models: a unified approach to scene parsing. Patt. Recogn. 84, 165–181 (2018)
- Alinia, P., Razzaghi, P.: Similarity based context for nonparametric scene parsing. In: 2017 Iranian Conference on Electrical Engineering (ICEE). IEEE, pp. 1509–1514 (2017)
- Taghanaki, S.A., Zheng, Y., Zhou, S.K., Georgescu, B., Sharma, P., Xu, D., Comaniciu, D., Hamarneh, G.: Combo loss: handling input and output imbalance in multi-organ segmentation. In: Computerized Medical Imaging and Graphics, vol. 75, pp.24–33 (2019)

- Moghbel, M., Ooi, C.Y., Ismail, N., Hau, Y.W., Memari, N.: A review of breast boundary and pectoral muscle segmentation methods in computer-aided detection/diagnosis of breast mammography. Artif. Intell. Rev. 53, 1873–1918 (2020)
- Mustra, M., Grgic, M.: Robust automatic breast and pectoral muscle segmentation from scanned mammograms. Signal Process. 93(10), 2817–2827 (2013)
- Subashini, T.S., Ramalingam, V., Palanivel, S.: Pectoral muscle removal and detection of masses in digital mammogram using CCL. Int. J. Comput. Appl. 1(6), 71–76 (2010)
- Tayel, M., Mohsen, A.: Breast boarder boundaries extraction using statistical properties of mammogram. In: IEEE 10th International Conference on Signal Processing Proceedings. IEEE, pp. 2468–2471 (2010)
- Czaplicka, K., Włodarczyk, J.: Automatic breast-line and pectoral muscle segmentation. Schedae Informaticae 20 (2011)
- Ergin, S., Esener, İ.I., Yüksel, T.: A genuine GLCM-based feature extraction for breast tissue classification on mammograms. Int. J. Intell. Syst. Appl. Eng, 4(91):124–129 (2016)
- 49. Hazarika, M., Mahanta, L.B.: A novel region growing based method to remove pectoral muscle from MLO mammogram images. In: Advances in Electronics, Communication and Computing: ETAEERE-2016, pp. 307-316. Springer Singapore (2018)
- Taghanaki, S.A., Liu, Y., Miles, B., Hamarneh, G.: Geometrybased pectoral muscle segmentation from MLO mammogram views. IEEE Trans. Biomed. Eng. 64(11), 2662–2671 (2017)
- Shrivastava, A., Chaudhary, A., Kulshreshtha, D., Singh, V.P., Srivastava, R.: Automated digital mammogram segmentation using dispersed region growing and sliding window algorithm. In: 2017 2nd international conference on image, vision and computing (ICIVC), pp. 366–370. IEEE (2017)
- Chen, Z., Zwiggelaar, R.: A combined method for automatic identification of the breast boundary in mammograms. In: 2012 5th International Conference on BioMedical Engineering and Informatics. IEEE, pp. 121–125 (2012)
- Nagi, J., Kareem, SA, Nagi, F, Ahmed, SK: Automated breast profile segmentation for ROI detection using digital mammograms. In: 2010 IEEE EMBS conference on biomedical engineering and sciences. IEEE (IECBES), pp. 87–92 (2010)
- Maitra, I.K., Nag, S., Bandyopadhyay, S.K.: Technique for preprocessing of digital mammogram. Comput. Methods Programs Biomed. 107(2), 175–188 (2012)
- Camilus, S.K., Govindan, V.K., Sathidevi, P.: Computer-aided identification of the pectoral muscle in digitized mammograms. J. Digital Imag. 23, 562-580 (2010)
- Xu, W., Li, L., Liu, W.: A novel pectoral muscle segmentation algorithm based on polyline fitting and elastic thread approaching. In: 2007 1st international conference on bioinformatics and biomedical engineering. IEEE, pp. 837–840 (2007)
- Shen, R., Yan, K., Xiao, F., Chang, J., Jiang, C., Zhou, K.: Automatic pectoral muscle region segmentation in mammograms using genetic algorithm and morphological selection. J. Digit. Imag. 31, 680–691 (2018)
- Slavković, I.M., Gavrovska, A., Milivojević, M., Reljin, I., Reljin, B.: Breast region segmentation and pectoral muscle removal in mammograms. Telfor J. 8(1), 50–55 (2016)
- Yin, K., Yan, S., Song, C., Zheng, B.: A robust method for segmenting pectoral muscle in mediolateral oblique (MLO) mammograms. Int. J. Comput. Assist. Radiol. Surg. 14, 237–248 (2019)
- Rampun, A., Morrow, P.J., Scotney, B.W., Winder, J.: Fully automated breast boundary and pectoral muscle segmentation in mammograms. Artif. Intell. Med. 79, 28–41 (2017)
- Akande, H.J., Olafimihan, B.B., Oyinloye, O.I.: Mammographic parenchymal patterns in asymptomatic women. Saudi J. Med. Med. Sci. 5(3), 232–237 (2017)

- Vilmun, B.M., Vejborg, I., Lynge, E., Lillholm, M., Nielsen, M., Nielsen, M.B., Carlsen, J.F.: Impact of adding breast density to breast cancer risk models: a systematic review. Eur. J. Radiol. 127, 109019 (2020)
- Lehman, C.D., Yala, A., Schuster, T., Dontchos, B., Bahl, M., Swanson, K., Barzilay, R.: Mammographic breast density assessment using deep learning: clinical implementation. Radiology 290(1), 52–58 (2019)
- Saffari, N., Rashwan, H.A., Abdel-Nasser, M., Singh, V.K., Arenas, M., Mangina, E., Herrera, B., Puig, D.: Fully automated breast density segmentation and classification using deep learning. Diagnostics 10(11), 988 (2020)
- Isola, P., Zhu, J.-Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1125–1134 (2017)
- Nickson, C., Arzhaeva, Y., Aitken, Z., Elgindy, T., Buckley, M., Li, M., English, D.R., Kavanagh, A.M.: AutoDensity: an automated method to measure mammographic breast density that predicts breast cancer risk and screening outcomes. Breast Cancer Res. 15, 1–12 (2013)
- Rouhi, R., Jafari, M., Kasaei, S., Keshavarzian, P.: Benign and malignant breast tumors classification based on region growing and CNN segmentation. Expert Syst. Appl. 42(3), 990–1002 (2015)
- Keller, B.M., Nathan, D.L., Wang, Y., Zheng, Y., Gee, J.C., Conant, E.F., Kontos, D.: Estimation of breast percent density in raw and processed full field digital mammography images via adaptive fuzzy c-means clustering and support vector machine segmentation. Med. Phys. 39(8), 4903–4917 (2012)
- Kallenberg, M., Petersen, K., Nielsen, M., Ng, A.Y., Diao, P., Igel, C., Vachon, C.M., et al.: Unsupervised deep learning applied to breast density segmentation and mammographic risk scoring. IEEE Trans. Med. Imaging 35(5), 1322–1331 (2016)
- Berg, W.A., Rafferty, E.A., Friedewald, S.M., Hruska, C.B., Rahbar, H.: Screening algorithms in dense breasts: AJR expert panel narrative review. Am. J. Roentgenol. 216(2), 275–294 (2021)
- Youk, J.H., Gweon, H.M., Son, E.J., Kim, J.-A.: Automated volumetric breast density measurements in the era of the BI-RADS fifth edition: a comparison with visual assessment. Am. J. Roentgenol. 206(5), 1056–1062 (2016)
- 72. Aliniya, P., Nicolescu, M., Nicolescu, M., George B.: Supervised pectoral muscle removal in mammography images. In: 22nd International conference of artificial intelligence in medicine

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Parvaneh Aliniya is a Ph.D. candidate in Computer Science at the University of Nevada, Reno. She earned her Master's degree from UNR in 2024. Prior to that, she obtained a Bachelor's in Information Technology Engineering and a Master's in Intelligent Systems Design from the Institute for Advanced Studies in Basic Sciences (IASBS). Her research focuses on computer vision and artificial intelligence, with an emphasis on applications in the medical domain.

Mircea Nicolescu is a Professor of Computer Science and Engineering at the University of Nevada, Reno and co-director of the UNR Computer Vision Laboratory. He received a PhD degree from the University of Southern California in 2003, a MS degree from USC in 1999, and a BS degree from the Polytechnic University Bucharest, Romania in 1995, all in Computer Science. His research interests include visual motion analysis, perceptual organization, vision-based surveillance, and activity recognition. Dr. Nicolescu's research has been funded by the Department of Homeland Security, the Office of Naval Research, the National Science Foundation and NASA. He is a member of the IEEE Computer Society.

Dr. Monica Nicolescu is a Professor with the Computer Science and Engineering Department at the University of Nevada, Reno and is the Director of the UNR Robotics Research Lab. Dr. Nicolescu earned her PhD degree in Computer Science from the University of Southern California (2003) at the Center for Robotics and Embedded Systems. She obtained her MS degree in Computer Science from USC (1999) and a BS in Computer Science at the Polytechnic University Bucharest (Romania, 1995). Her research interests are in the areas of humanrobot interaction, robot control, learning, and multi-robot systems. Dr. Nicolescu's research has been supported by the National Science Foundation, the Office of Naval Research, the Department of Energy and Nevada Nanotech Systems. In 2006 she was a recipient of the NSF Early Career Development Award (CAREER) Award for her work on robot learning by demonstration.

George Bebis received the B.S. degree in Mathematics and the M.S. degree in Computer Science from the University of Crete, Greece, in 1987 and 1991, respectively, and the Ph.D. degree in Electrical and Computer Engineering from the University of Central Florida, Orlando, in 1996. Currently, he is a Foundation Professor in the Department of Computer Science and Engineering (CSE) at the University of Nevada, Reno (UNR) and Director of the Computer Vision Laboratory (CVL). From 2013 to 2018, he served as Department Chair of CSE-UNR. His research interests include computer vision, image processing, pattern recognition, machine learning, and evolutionary computing. His research has been funded by NSF, NASA, ONR, NIJ, and Ford Motor Company. Dr. Bebis is an Associate Editor of the Machine Vision and Applications Journal and serves on the editorial board of the International Journal on Artificial Intelligence Tools, and the Computer Methods in Biomechanics and Biomedical Engineering: Imaging and Visualization journal. He has served on the program committees of various national and international conferences and is the founder/main organizer of the International Symposium on Visual Computing (ISVC) and the International Symposium on Mathematical and Computational Oncology (ISMCO).