ASP Loss: Adaptive Sample-Level Prioritizing Loss for Mass Segmentation on Whole Mammography Images

Parvaneh Aliniya*(*aliniya@nevada.unr.edu*)[1][0009−0005−0925−743X], Mircea Nicolescu[1][0009−0008−3748−3559], Monica Nicolescu[1][0009−0009−8748−7918], and George Bebis[1][0000−0003−0966−6063]

University of Nevada, Reno, NV 89557, USA

Abstract. Alarming statistics on the mortality rate for breast cancer are a clear indicator of the significance of computer vision tasks related to cancer identification. In this study, we focus on mass segmentation, which is a crucial task for cancer identification as it preserves critical properties of the mass, such as shape and size, vital for identification tasks. While achieving promising results, existing approaches are mostly hindered by pixel class imbalance and various mass sizes that are inherent properties of masses in mammography images. We propose to alleviate this limitation on segmentation methods via a novel modification of the common hybrid loss, which is a weighted sum of the cross entropy and dice loss. The proposed loss, termed Adaptive Sample-Level Prioritizing (ASP) loss, leverages the higher-level information presented in the segmentation mask for customizing the loss for every sample, to prioritize the contribution of each loss term accordingly. As one of the variations of U-Net, AU-Net is selected as the baseline approach for the evaluation of the proposed loss. The ASP loss could be integrated with other existing mass segmentation approaches to enhance their performance by providing them with the ability to address the problems associated with the pixel class imbalance and diverse mass sizes specific to the domain of breast mass segmentation. We tested our method on two publicly available datasets, INbreast and CBIS-DDSM. The results of our experiments validate our approach and show a significant boost in the performance of the baseline method while outperforming state-of-the-art mass segmentation methods.

1 Introduction

Despite the significant progress in breast cancer screening in recent decades, breast cancer has been continuously one of the cancer types with the highest mortality rate among women [1]. Automated breast cancer detection could alleviate this problem in different ways - for example, when used to reduce the cost of a second reader [2, 3], especially when developed for mammography input images, which are some of the most common screening tools with reported effectiveness in reducing mortality rate [4].
Recently there has been a significant increase in the accuracy of identifying different abnormalities in breast tissues, thanks to deploying and tailoring deep learning approaches for each task in this domain. However, the performance of these methods is restricted by one inherent data-specific challenge for the majority of approaches on the whole mammogram: pixel class imbalance in which one class is underrepresented [5]. In this case, it could be considered as a very small size of the foreground (abnormality) as compared to the background (combination of the image background and other breast regions). This limitation, coupled with various sizes of masses, ranging from

micro-masses to relatively large ones, constrains the performance of existing methods. In light of this observation, in this work, we propose customizing the hybrid loss function for each sample through the Adaptive Sample-Level Prioritizing (ASP) loss, which significantly outperforms the common setting for the hybrid loss function used in the existing mass segmentation methods for mammography images.

The ASP loss functions proposed in this paper are composed of two losses: cross entropy and dice losses [5]. Instead of fixed weights [6, 7] for the loss terms, we propose to use the ratio of the mass to the image size as an indicator of the severity of pixel class imbalance and the differentiating factor for various mass sizes. This allows for balancing the contribution of each loss with respect to each sample in the training process. To achieve this goal, we propose three strategies for prioritizing one of the loss terms over another adaptively, based on the ratio of the mass, which is a relevant feature of each sample to the weighting of the loss terms.

As the first variation, we explore the idea of a quantile-based strategy, which accentuates the contribution of one loss over the other based on the quantile to which the sample belongs. To take the distribution of the samples based on the ratio into consideration while prioritizing the loss terms, a second strategy is a cluster-based prioritizing approach which focuses more on the proximity of the samples rather than the quantities for grouping. Lastly, we explore the idea of parameterizing the ASP loss based on learning the mapping from the ground truth segmentation for each sample to weights for loss terms without explicitly incorporating the ratio into the ASP. All three variations of ASP have been tested on two benchmark datasets for mass segmentation: INbreast [8], and CBIS-DDSM [9]. As the baseline method, AU-Net[6], which is a modern and effective variation of U-Net [10] has been chosen. The results of our experiments illustrate that ASP loss has a significant effect on the performance of the method.

The contributions of this paper are the following:

- Proposing an adaptive sample-level prioritizing (ASP) loss function for mass segmentation on whole mammograms.
- Developing three different prioritizing strategies for the proposed ASP loss: quantile-based, cluster-based, and learning-based priority.
- Evaluating the proposed method with all its variations on two benchmark datasets, INbreast and CBIS-DDSM.
- Quantitatively analyzing and comparing the findings from our experimental results for the proposed method with state-of-the-art approaches.

In the following, first, the related work is presented, then the proposed method is delineated in detail, and finally, our experimental results (including the ablation study and the comparison with state-of-the-art methods) are analyzed from various perspectives.

2 Related Work

Recently, deep learning-based approaches have shown great promise in abnormality identification in medical images, with several studies proposing various methods to achieve more accurate mass segmentation. In this section, we seek to provide a brief review of related work in deep learning-based approaches for breast mass segmentation, categorized into two groups related to the main focus of this study: breast mass segmentation on whole mammograms and loss functions mainly proposed for binary segmentation of medical images.

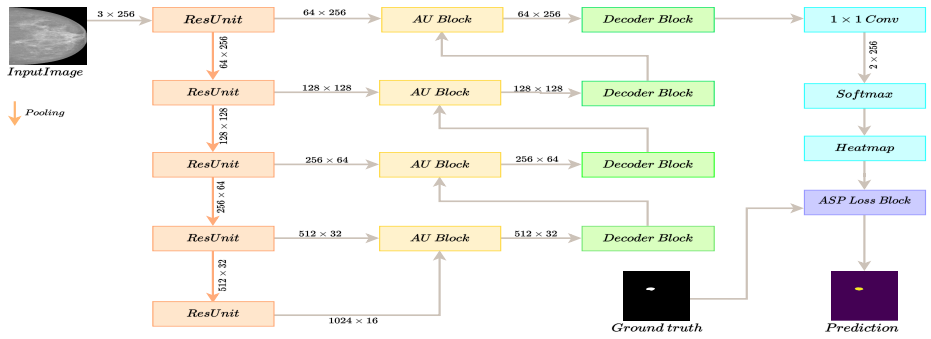
2.1 Mass Segmentation on Whole Mammograms

In general, breast mass segmentation could be categorized into two subsets based on the input to be used: region of interest (RoI) based methods, and approaches using whole mammograms. Approaches proposed for RoI of mass [11] have different challenges and strategies to address them compared to methods using whole mammogram images [6, 7]. For instance, the severity of the pixel class imbalance and diverse mass sizes is less significant in RoIs, as the majority of the background is removed. Therefore in this section, we focus on the approaches using the whole mammogram.
Inspired by [12], which is among the pioneer deep learning-based approaches for semantic segmentation, U-Net [10] proposes a fully convolutional symmetric encoder-decoder architecture that is specifically useful for segmentation tasks when there is limited data available, that makes it favorable for medical images in which data scarcity is a common limitation. U-Net combines high-level semantic information from the decoder path with low-level location information from the encoder. [13] proposes a similar encoder-decoder architecture (with Dense Blocks), in which the authors introduced multiscale information to the network by leveraging the idea of using the results of atrous convolution [14] with various sample rates for the last encoder block, to enhance the performance of the segmentation network without additional parameters. [15] is another U-Net-based approach that uses a densely-connected network in the encoder, and for the decoder, CNN is used with attention gates. Another research work in the scope of multi-scale studies is [16], where an adversarial framework incorporates the idea of using multiple networks for different scales for the discriminator, and an improved version of U-Net is used for segmentation (generator). In [17], the authors use the error of the outputs of intermediate layers in comparison with the ground truth labels as a supervision signal to boost the performance of the model.
In [6], authors propose an attention-guided dense-up-sampling asymmetric encoder-decoder network that has an intermediate up-sampling block with a channel-wise attention mechanism designed to use the useful information presented in both low and high-level features. With the goal of addressing the low performance of the U-Net on small-size masses, [7] proposes using a selective receptive field module which has two parts, one for generating several receptive fields with different sizes and one for selecting the appropriate size of the receptive fields. Our work is closely related to [6] and [7].

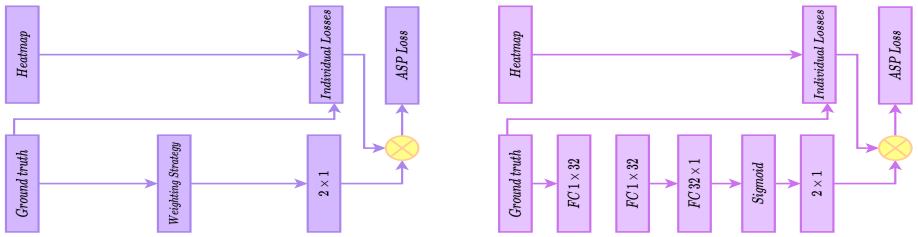
2.2 Loss for Segmentation in Medical Imaging

As the mathematical representation of the objective of a deep learning-based approach, the loss function has a significant impact on the performance of the method. Hence the choice of a loss function capable of appropriately reflecting the objective will lead to a large boost in the learning of the network in the segmentation task. This section aims to provide a concise summary of related loss functions for segmentation, specifically for medical imaging. Pixel class imbalance is one of the main challenges for medical image segmentation due to the fact that a majority of pixels belong to the background or normal regions. This limitation, alongside various mass sizes, makes the formulation of the appropriate loss function challenging.
Binary Cross Entropy (BCE) [18] for the segmentation task (Equation 1) is defined as the classification of pixels to the positive (foreground) and negative (background) classes. Weighted Binary Cross Entropy [19] and Balanced Cross Entropy [20] are variants of BCE that differentiate between the effect of false positives and false negatives factors. The focal loss [21] could be considered as another variation of BCE in which the

magnitude of the loss changes according to the hardness of the example based on the confidence of the model in false prediction. Dice loss is a commonly used loss function in the presence of pixel class imbalance [22], which is the ratio of the number of correctly classified pixels to the total number of real and predicted positive pixels (Equation 2). Tversky loss [23] aims to accentuate the contribution of the false positives and the false negatives terms in the dice loss via weighting these terms.
Combo loss [24] combines the dice and modified BCE in which the contribution of false positive and false negative has been controlled via a weighting strategy. Various compound losses have been proposed [21, 23, 24] to use the benefit of different losses. For instance, in the Combo loss [24], dice loss is beneficial for handling pixel class imbalance problems and the BCE for smoother training. Our proposed method is similar to this category in terms of using dice and BCE losses but differs in terms of the weighting strategy, in which the ratio of the positive class has been utilized as a weighting strategy between two loss terms (dice and BCE losses). In addition, instead of directly using the ratio, we use it for grouping samples.



(a) The overall architecture of the method.



(b) Quantile and cluster based ASP Block. (c) Learning-based ASP Block.

Fig. 1: An overview of the proposed method.

3 Methodology

In this section, first, we provide an overview of hybrid loss which is one of the common loss functions for mass segmentation methods with mammography input type. The hybrid loss is defined as a weighted sum of cross entropy and dice loss, which are defined as follows:

$$L_{BCE} = -\Big(ylog(\hat{y}) + (1-y)log(1-\hat{y})\Big) \tag{1}$$

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

$$L_H = \alpha L_{Dice} + \beta L_{BCE} \tag{3}$$

Here y and $\hat{y}$ are the ground truth and the predicted segmentation maps. α and β are the weighting parameters in the hybrid loss denoted as L_H in Equation 3. While the cross entropy loss (Equation 1) is defined to incorporate notions of false positive and negative in the learning process, dice loss (Equation 2) specifically captures correctly classified, positive classes. Due to the presence of pixel class imbalance problem, the combination (Equation 3) of the two losses will provide a better learning signal. In addition, the cross entropy helps mitigate the unstable training of the dice loss when used alone [21, 22]. While improving performance compared to using only one of them, we speculate that a constant weighting strategy for all the samples, regardless of the severity of the pixel class imbalance problem and mass size specific to each sample, restricts the performance of the hybrid loss (Equation 3). We propose to incorporate the sample-specific signal for the hybrid loss weighting strategy. In the following, three different strategies proposed for the ASP loss are explained.

3.1 Adaptive Sample-Level Prioritizing Loss

Formally, given a training set of N images and the segmentation masks, the baseline method learns a mapping function from the input images to predicted segmentation results. In this study, AU-Net is selected as the baseline method, and the architecture for AU-Net is depicted in Figure 1a. For the encoder and the decoder, ResUnit and the basic decoder proposed in the AU-Net have been used. The details of the Attention-guided Up-sampling Block (AU Block) are presented in the AU-Net approach [6]. One relatively common and effective loss for breast mass segmentation is the hybrid loss function defined in Equation 3. However, one of the main challenges is to use an effective weighting strategy to capture the strength of both losses.
An intuitive idea would be to leverage the size of the masses as a data-driven signal to adaptively adjust the weighting parameters in lieu of relying on a static hyperparameter setting. To accomplish this, first the ratio of the masses to the image size is extracted for each training sample in a set $R = \{r^1, ..., r^N\}$.

$$r^i = \frac{\sum_{j=1}^{H\times W} y_j^i}{H \times W} \tag{4}$$

Here y is the ground truth label of the j^{th} pixel in the i^{th} sample. Now, the primary question is how to integrate this supplementary piece of information into the weighting scheme of the two loss functions. Based on the definition of binary cross entropy and the dice loss, the cross entropy loss is more favorable for larger mass sizes (as the size of the image is incorporated in the loss, BCE might penalize the mismatch between prediction and ground truth of the smaller masses less), while the dice loss is better suited for smaller masses. Therefore, we explore the idea of utilizing the correlation between mass size and loss type to define the weights for each loss term as defined in Equation 5.

$$L_{ASP}^i = (I_{Dice} + (1 - p^i)\gamma)L_{Dice}^i + (I_{BCE} + (p^i)\gamma)L_{BCE}^i \tag{5}$$

Here I_{Dice} and I_{BCE} are the initial weights of each loss term, and γ and p are the magnitude parameter and prioritizing variable for the ASP loss. While p selects the more relevant loss term to prioritize for the i^{th} sample, γ controls the amount of the effect of the loss term, which could be constant or sample-based. Also, γ could be different for both loss terms. It should be noted that I_{Dice} and I_{BCE} could correspond to α and β in the traditional weighting presented in Equation 3. In addition, index i in Equation 5 indicates that the proposed loss is a sample-level loss compared to the static loss in Equation 3. There are various ways to define p, which is the main component of ASP loss. We explore three different variations.

Quantile-based Strategy One basic idea would be to use the quantile of the values in R to divide the samples into two groups to prioritize different loss terms, as defined in Equation 6:

$$p_Q^i = \begin{cases} 0 & \mathrm{r}^i < Q_2 \\ 1 & \text{Otherwise} \end{cases} \tag{6}$$

Here Q_2 is the median of R and p_Q^i is the prioritizing variable for the i^{th} sample using the quantile-based approach. It should be noted that the other quantiles could be used. The number of samples in the two groups will be equal.

Cluster-based Strategy Due to the pixel class imbalance problem, as the distribution of the mass sizes is skewed toward smaller masses, the group with mass sizes above the median (in the second quantile) will still have diverse mass sizes. To alleviate this problem, K-means clustering with two centers has been used to cluster the masses based on the relative ratios, providing a more meaningful (in terms of relation with the mass size) division of the samples for the ASP loss. The definition of the cluster-based prioritizing variable is given in Equation 7:

$$p_C^i = \begin{cases} 0 & \mathrm{r}^i \in C_s \\ 1 & \text{Otherwise} \end{cases} \tag{7}$$

Here C_s is the center of the cluster for the smaller masses, and p_C^i denotes the prioritizing variable of the ASP loss using the cluster-based approach. Both quantile and cluster based strategies are defined in the way they use the ratio as a guiding signal for learning. The ASP loss diagram for quantile and cluster based prioritizing variables has been shown in Figure 1b.

Learning-based Strategy In the last prioritizing strategy, the network learns the weights for the loss terms through the learning process. The ASP weight learning module is depicted in Figure 1c and formulated in Equation 8. This module (denoted as f in Equation 8) takes the network prediction and the ground truth segmentation mask as inputs and generates the weights for the ASP loss. f consists of one convolution layer, two fully connected layers, and a sigmoid activation function. The indexes of 1 and 2 in the learning-based ASP loss (Equation 8) indicate the output index as the output is a 2×1 matrix. The motivation behind the learning-based prioritizing strategy is to let the network learn the connection between the segmentation masks and the loss terms, which could capture different connections beyond the mass ratio.

$$L_{ASP}^i = f(y, w)_1 L_{Dice}^i + f(y, w)_2 L_{BCE}^i \tag{8}$$

Here w is the parameters for the ASP loss in Figure 1c. As mentioned before, the proposed loss functions could be used in combination with different architectures for binary segmentation tasks for applications with inherent pixel class imbalance problems. In this study, AU-Net, as a new and broadly used variation of U-Net, has been chosen as the baseline architecture.

4 Experimental Results

Our findings on the two benchmark datasets suggest that the ASP strategies proposed in this study are effective in improving performance specifically for the varying size of the masses and handle the pixel class imbalance better than previous state-of-the-art methods, which is shown by boosted detection of highly imbalanced and diverse mass sizes. This section provides an overview of the selected datasets for evaluating the proposed method, along with definitions of evaluation metrics. Next, our findings through ablation studies and comparisons with previous state-the-art-methods are presented and analyzed in the following sections.

4.1 Datasets

For both datasets, images have been resized to 256×256 pixels. No data augmentation or pre-training was used. The validation set was utilized to prevent overfitting by hyperparameter tuning for the number of epochs. The number of epochs is set to a value between 170 and 240, and the batch size is four based on the setting of the baseline method. The learning rate is initially set to 10-e4, and the step decay policy with a decay factor of 0.5 has been used for experiments conducted on both datasets. It should be noted that all the mass images have been utilized in our experiments regardless of their category (benign or malignant).

INbreast Dataset A group of 150 cases with a total number of 410 images (some of the cases have several images for different views and follow-ups) are used in the INbreast dataset. 107 of the images contain masses (the total number of masses is 116) which have been used in this study. Image enhancement was not used for the INbreast dataset. For the validation, due to limited samples, 5-fold cross-validation is used with the random division of 80%, 10%, and 10% for train, validation, and test sets, respectively.

CBIS-DDSM Dataset CBIS-DDSM has a total of 1944 cases, 1591 of which contain masses. The standard split for train and test (1231 and 360 images for train and test sets, respectively) is used in this study. For the validation set 10% of the training set is randomly sampled. Before using the samples in CBIS-DDSM, the artifacts were removed, and images were cropped and resized.

4.2 Evaluation Metrics

For evaluating the performance of the proposed method for mass segmentation, the Dice Similarity Coefficient (DSC), Relative Area Difference ΔA, Sensitivity, and Accuracy have been selected due to the complementary information that they provide. For the breast mass segmentation task, as the masses occupy a trivial portion of the image, using accuracy solely would not reflect the performance of the methods accurately. Hence, using additional metrics such as sensitivity that captures the percentage of the positive class in the ground truth that has been correctly predicted will be helpful. While

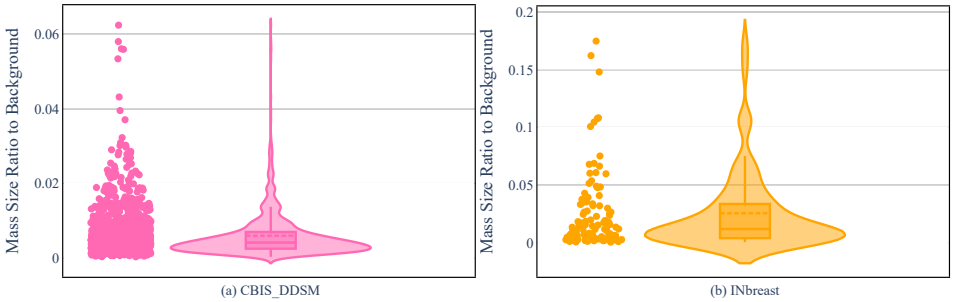


Fig. 2: Violin plots for the CBIS-DDSM (a) and INbreast (b) based on the mass ratio.

providing more meaningful insight into the performance of the method for the positive class, sensitivity lacks the ability to incorporate the false positive rate in the calculations. DSC, which measures the ratio of the correctly predicted positive pixels over the number of positive areas in both ground truth and the prediction mask, considers the false positive rate in the calculations. One other informative measurement instrumental for the evaluation of the proposed methods is ΔA which measures the difference in the sizes of the actual and predicted masses (a smaller value for ΔA indicates closer sizes for masses).

4.3 Comparison of Dataset Characteristics

In this section, we aim to specifically analyze the difference in the mass ratios between the two datasets. To this end, the distributions of masses in the datasets have been depicted in Figure 2 using violin plots which are a combination of box and kernel density plots to capture both summary of statistics and the density of the variables. The length of the line above the box shows 4^{th} quartile, the length of the box represents interquartile (2^{th} *and* 3^{th} quartiles), and the length of the line below the box represents the 1^{th} quartile. The dashed line inside the box is the mean, and the solid line is the median. At one glance, it is clear that the CBIS-DDSM (Figure 2a) has a smaller range compared to the INbreast dataset (Figure 2b). Also, the length of the interquartile range is smaller for the CBIS-DDSM dataset, indicating less diversity in the ratio for the majority of the masses. Also, closer mean and median in the CBIS-DDSM dataset suggest less skewed data. These statistics will be helpful in the analyzing section for the results of the proposed method on the two datasets.

4.4 Comparison with State-of-the-art Methods

For comparison purposes, AU-Net (baseline) and ARF-Net, which are two state-of-the-art methods, have been selected. We used the official implementation of the AU-Net, and the setting described in the AU-Net paper [6] (only the network was publicly available). ARF-Net is one of the state-of-the-art methods for mass segmentation on whole mammograms, which is related to our study as it incorporates the idea of designing the network in a way that leverages different sizes. For the ARF-Net, as the implementation was not publicly available, the method was implemented to the best of our understanding based on the paper. As we did not use any pre-training or data augmentation, for the purpose of a better comparison, these models also have been trained in

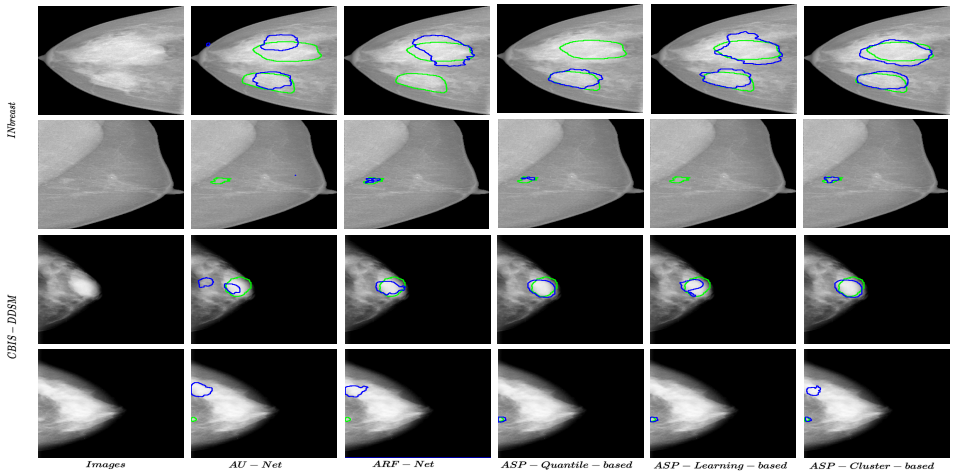


Fig. 3: Results for three variations of the ASP, AU-Net, and ARF-Net approaches.

the same manner. For the quantile-based strategy, three hyperparameters are needed. For I_{Dice} and I_{BCE}, 0.125 and 0.25 have been selected in this study for INbreast and CBIS-DDSM, respectively. For γ, a value between [0.25, 0,35] has been used with different values for dice and cross entropy terms. For the cluster-based strategy, the number of centers is a hyperparameter, for which 2 was selected. These hyperparameters have been selected experimentally.

Experimental Results for INbreast Dataset The results of the experiments of the trained model on INbreast are presented in Table 1. As shown in bold, the best results have been achieved by the proposed method. While showing a variety in the results for different prioritizing strategies in the ASP loss, all of them improved on the AU-Net (baseline) method and surpassed the ARF-Net approach, which incorporates the idea of using different sizes in the architecture. The largest improvement has been achieved using the cluster-based prioritizing strategy for the ASP loss (DSC: +8.86%, ΔA: -4.4%, Sensitivity: + 9.26%, Accuracy: +0.32 %), which is consistent and significant across all metrics. We attribute this to the fact that the cluster-based approach captures the distribution of the data in terms of the ratio better than other strategies. As shown in Figure 2b, the smaller ratios are mostly in the same area in the interquartile range, and bigger ratios are at different ranges. This makes the cluster-based approach more appropriate for grouping the data as it represents the real distribution more accurately compared to the quantile-based approach, which uses the median as a factor of division. As for one of the groups (values above the median), there are a variety of sizes, from small sizes to larger ones. Some of the small masses will prioritize the loss terms in the same manner as the larger masses. We believe this might contribute to the smaller improvement in the quantile-based approach for the INbreast dataset.

The performance of the learning-based strategy is also superior compared to previous approaches. It confirms the idea that a learning-based strategy captures relevant properties of the mass in the ground truth for weighting the losses. Figure 3 shows that the learning-based ASP has superior performance in most of the cases compared to the previous methods.

Table 1: Comparison of ASP with state-of-the-art approaches for INbreast.

Method	DCS	ΔA	Sensitivity	Accuracy
ARF-Net	70.05	30.37	59.59	98.71
AU-Net	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**

Figure 3 (first two rows) shows the results for some images in the test set for all ASP versions, AU-Net, and ARF-Net for INbreast dataset. The examples have been selected in such a way that they cover different sizes to showcase the prediction power of the methods on both types of masses according to the mass ratio. The green lines are the contours of the ground truth masks, and the blue lines are the contours of the prediction masks. As shown, all ASP variations predict masses in both sizes better than state-of-the-art methods. For masses with large size (larger ratio), the cluster-based and learning-based strategies outperform AU-Net and ARF-Net clearly, and the cluster-based approach has a more accurate contour. The quantile-based method has comparable performance. This might stem from the aforementioned argument that based on the distribution of the data in INbreast, the quantile-based approach does not capture the differences thoroughly like the cluster-based approach, which is the variation with the best overall performance. For the small ratio, the quantile-based and cluster-based methods have better performance compared to previous techniques.

Table 2: Comparison of ASP with state-of-the-art approaches for CBIS-DDSM.

Method	DCS	ΔA	Sensitivity	Accuracy
ARF-Net	48.82	11.47	47.27	99.43
AU-Net	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

Experimental Results for CBIS-DDSM Dataset The ASP loss also has superior performance on the CBIS-DDSM dataset compared to AU-Net and ARF-Net approaches presented in Table 2. Quantile-based ASP has the best performance on CBIS-DDSM, which is different from the results in the INbreast dataset. We speculate that the reason stems from the difference between the distribution in the two datasets as CBIS-DDSM is less skewed regarding the ratios. While cluster-based ASP divides the data for INbreast better compared to quantile-based ASP, for the CBIS-DDSM dataset, as the data is mostly concentrated in the smaller mass size, the cluster-based strategy will place most of the data in one cluster. Therefore, the weighting strategy will not capture the differences accurately. On the other hand, quantile-based has a better data division as it uses the median. Even though there will be a range of small to large size masses in the second group, but as a small portion of the samples corresponds to large ratios, most of the samples will be weighted correctly compared to the cluster-based strategy, in which a large portion of the samples are in the same group which diminishes the effect of grouping. Figure 3 (last two rows) shows that all three ASP strategies have better performance on both large and small mass sizes when compared to the previous

methods. Specifically, the quantile-based approach has the most accurate contours when compared to the two other variations and previous methods.

5 Conclusion

In this study, in order to use available additional information in the segmentation mask for mass segmentation in whole-view mammography images, we propose to modify the hybrid loss weighting strategy adaptively based on the mass size in each sample. Prioritizing loss terms in hybrid loss based on the property of the masses (size) in the sample-level is the main novel contribution of our work. Three different variations of the ASP loss based on different weighting strategies have been proposed in this study. The quantile-based strategy aims to group the data based on size while maintaining the same number of samples in both groups. The cluster-based approach takes the distribution of the data into consideration while dividing the data into groups. Finally, the learning-based strategy learns the prioritizing weights during the training process, so that in the beginning it is random, and during training it will learn the appropriate weights, based on the segmentation mask. The main idea for this strategy is to generate the weights nonlinearly according to the segmentation mask, which will incorporate more information into the learning of the weights. Our experimental results demonstrate a substantial improvement based on commonly used evaluation metrics on benchmark datasets: INbreast and CBIS-DDSM. Considering the mass sizes in individual samples by adaptively revising the hybrid loss allows the proposed loss to achieve superior performance. In future works, we plan to formulate the loss to incorporate additional information in the prediction of the network alongside the ground truth labels.

References

1. Siegel, R.L., Miller, K.D., Fuchs, H.E., Jemal, A.: Cancer statistics, 2022. CA: a cancer journal for clinicians, 72(1),7-33, (2022).
2. 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).
3. 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).
4. 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).
5. 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 Networks, 25, 141-145, (2012).
6. 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. Physics in Medicine and Biology, vol. 65(5), p.055005, (2020).
7. 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. Biomedical Signal Processing and Control, 71, 103178, (2022).
8. Moreira, I.C., Amaral, I., Domingues, I., Cardoso, A., Cardoso, M.J., Cardoso, J.S.: Inbreast: toward a full-field digital mammographic database. Academic radiology, 19(2), 236-248, (2012).

9. 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. Scientific data, 4(1), 1-9, (2017).
10. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation.In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015, pp. 234-241. Springer International Publishing, (2015).
11. Baccouche, A., Garcia-Zapirain, B., Castillo Olea, C., Elmaghraby, A.S.: Connected-UNets: a deep learning architecture for breast mass segmentation. NPJ Breast Cancer, 7(1), 151, (2021).
12. 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).
13. 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. Journal of healthcare engineering, (2019).
14. 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 transactions on pattern analysis and machine intelligence, 40(4), 834-848, (2017).
15. 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).
16. 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).
17. Rajalakshmi, N.R., Vidhyapriya, R., Elango, N., Ramesh, N.: Deeply supervised u-net for mass segmentation in digital mammograms. International Journal of Imaging Systems and Technology, 31(1), 59-71, (2021).
18. 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, pp. 743-746. IEEE, (2004).
19. 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).
20. Xie, S., Tu, Z.: Holistically-nested edge detection. In: Proceedings of the IEEE international conference on computer vision, 1395-1403, (2015).
21. 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, Computerized Medical Imaging and Graphics, 95, 102026, (2022).
22. 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).
23. 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).
24. 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).