Journal of *Imaging*

MDPI

Article

Improved Loss Function for Mass Segmentation in Mammography Images Using Density and Mass Size

Parvaneh Aliniya *, Mircea Nicolescu *, Monica Nicolescu and George Bebis

Computer Science and Engineering Department, College of Engineering, University of Nevada, Reno, 89557 NV, USA; monica@unr.edu (M.N.); bebis@cse.unr.edu (G.B.)
* Correspondence: aliniya@nevada.unr.edu (P.A.); mircea@cse.unr.edu (M.N.)

Abstract: Mass segmentation is one of the fundamental tasks used when identifying breast cancer due to the comprehensive information it provides, including the location, size, and border of the masses. Despite significant improvement in the performance of the task, certain properties of the data, such as pixel class imbalance and the diverse appearance and sizes of masses, remain challenging. Recently, there has been a surge in articles proposing to address pixel class imbalance through the formulation of the loss function. While demonstrating an enhancement in performance, they mostly fail to address the problem comprehensively. In this paper, we propose a new perspective on the calculation of the loss that enables the binary segmentation loss to incorporate the sample-level information and region-level losses in a hybrid loss setting. We propose two variations of the loss to include mass size and density in the loss calculation. Also, we introduce a single loss variant using the idea of utilizing mass size and density to enhance focal loss. We tested the proposed method on benchmark datasets: CBIS-DDSM and INbreast. Our approach outperformed the baseline and state-of-the-art methods on both datasets.

Keywords: breast cancer; mass segmentation; loss function; density; adaptive sample-level prioritizing loss; hybrid loss

Citation: Aliniya, P.; Nicolescu, M.; Nicolescu, M.; Bebis, G. Improved Loss Function for Mass Segmentation in Mammography Images Using Density and Mass Size. *J. Imaging* **2024**, *10*, 20. https://doi.org/10.3390/jimaging10010020

Academic Editor: Antoine Vacavant

Received: 28 November 2023
Revised: 31 December 2023
Accepted: 4 January 2024
Published: 9 January 2024

1. Introduction

Breast cancer is one the most common cancer types among females [1], and has been one of the main contributors to the high mortality statistics of cancer in females. The early detection of cancer has a significant impact on the improvement of the five-year survival rate compared with late-stage cancer [2]. Thanks to the cost-effectiveness and availability of mammography as a screening tool, it is one of the main tools for the early identification and treatment of the disease.

In recent years, the ability of deep-learning-based methods in automated feature extraction has enhanced the performance of automated breast cancer identification, despite limitations such as data scarcity in this domain. However, reading the mammography for findings has specific challenges, such as misdiagnosis or high cost of second readers. Therefore, using computer-aided diagnosis (CAD) systems could be highly beneficial to radiologists and patients.

Initially, CAD systems were designed based on traditional machine learning approaches [3–7]; however, the excellent performance of deep-learning-based approaches [8–11] in recent years has encouraged the research community in the medical domain to utilize and customize them to the special needs in this domain [12]. Despite the reported effectiveness of using deep learning methods for the identification of breast cancer [13–15], there are several challenges, including pixel class imbalance and various breast densities constraining the performance of the methods. Addressing these problems could greatly enhance the performance of the methods. There are various abnormalities in the mammography images, including masses, asymmetrical breast tissue, micro-calcification, and

architectural distortion of breast tissue [16]. Among the abnormality types, masses are reportedly major contributors to breast cancer [16]. Therefore, in this paper, we aim to address the identification of masses that generally has two forms: mass detection [17–20] and mass segmentation [21–25]. Mass segmentation provides more comprehensive information, including border information; hence, in this paper, the main task is mass segmentation.

Generally, deep-learning-based methods are the current state-of-the-art in most computer vision tasks. Among the various factors affecting the performance of deep-learning methods, the architecture of the network and loss function are of vital importance. A majority of the proposed methods aim to improve and customize the architecture of the network [21–26], which has proven to be effective. Another line of research attracting major attention from the research community is the customization of the loss function. As the loss function presents the objective of the network, adjustments to the definition of the loss could improve the training [27–31]. Loss functions for segmentation, in general, compare the pixels in the prediction and ground truth masks. While some loss functions consider pixels as individual entities (pixel-level losses) [30,32], others consider the neighboring pixels when calculating the loss for a central pixel (region-level losses) [33,34]. This, in turn, allows the loss function to incorporate the dependencies between the pixels in the loss calculation. Pixel-level losses mostly differ in the way they treat the TP, TN, FP, and FN rates. A common trend when using the proposed losses for mass segmentation is to use hybrid losses, which constitute a weighted sum of two or more losses.

While they are effective, there are several aspects for improvement in the currently available losses. Firstly, they neglect to use the extra information available in the dataset in the training to improve the performance. Secondly, the definition of the loss function is static; for example, the weights for the losses in the hybrid setting are fixed, despite the fact that the severity of a problem, such as the pixel class imbalance or the density, can vary across samples.

In this paper, we propose novel loss functions, taking the aforementioned shortcomings into consideration to improve the performance of the method in the presence of pixel class imbalance and higher densities. To this end, we propose two loss categories that use extra information in the data to address the problems in different ways. While they utilize different strategies, one aspect they share is being sample-level, meaning that the formulation of the losses depends on the information in each sample. In addition, both approaches use extra information to prioritize the more suitable loss term for each sample in the hybrid loss. Therefore, the general name for these losses is "Adaptive Sample-Level Prioritizing" (ASP) losses.

The first loss function, which utilizes the ratio of the mass in each sample for re-weighting the loss term in a hybrid setting is referred to as R-ASP loss (R stands for the ratio), as it uses the ratio of the mass size to the image size for prioritizing the loss terms over each other. The second loss, named D-ASP (D stands for density), is inspired by the same idea, but differs with respect to the information it uses for the re-weighting and the losses incorporated into the hybrid loss. Both losses have been tested on benchmark datasets, the Curated Breast Imaging Subset of DDSM (CBIS-DDSM) and INbreast, using AU-Net as the baseline architecture. In addition, we propose customizing the ASP method for the selection of the focusing parameter in focal loss to examine the potential of ASP for a single loss. The results of the experiments demonstrate a robust improvement in the performance of the method to varying degrees across the proposed ASP losses, outperforming the state-of-the-art methods.

The contributions of this paper are as follows:

- Proposing a general framework for adaptive sample-level prioritizing losses.
- Introducing two variations of ASP loss that alleviate the limitation of performance caused by pixel class imbalance and density categories.
- Customizing focal loss to use the ratio and density for the selection of the focusing parameter.

- Evaluating the proposed losses on CBIS-DDSM [35] and INbreast [16] datasets using AU-Net architecture.
- Performing ablation study on INbreast to investigate the impact of different parameters.
- Comparing the ASP losses with traditional hybrid loss and state-of-the-art methods.

In the following sections, a comprehensive review of the related work is provided, followed by the proposed method and a detailed explanation of each of the losses. Next, the experimental settings for testing the proposed losses, as well as an analysis of the results, is provided. Finally, the paper ends with Sections 5 and 6.

2. Related Work

When it comes to the outstanding performance gained by the adoption of deep learning as an end-to-end approach for automated feature extraction and segmentation, mass segmentation is no exception. As a result of the transformation in recent years, the majority of the proposed methods fall into the deep-learning category, which is the main target of this paper. Therefore, in this section, mass segmentation approaches [36–42] using deep-learning-based methods are first reviewed. In addition, as the main contribution of the paper is to propose new ways to incorporate additional information in the hybrid loss setting for binary segmentation, reviewing the losses for this task is also essential; hence, the second subsection is devoted to reviewing the binary segmentation losses, mostly focusing on the ones proposed for medical applications.

2.1. Review of Mass Segmentation Approaches

Supervised deep-learning-based mass segmentation methods could be performed on the region of interest (ROI) [43,44] of the masses or the full field of view (FOV) (or entire mammography image). As the main task in this paper was to perform segmentation on the image, only the previous work for mass segmentation on the whole view mammography images was reviewed in this section. This was because the proposed losses were for alleviating problems such as the pixel class imbalance that was presented in the whole-view mammography image.

While the authors of [45,46] were among the very first to design a segmentation method using deep learning, U-Net [12] is a pioneering work in the realm of deep-learning-based segmentation for medical applications that is still inspiring new methods to this day.

The fully convolutional network (FCN) [45] introduced a network with skip connections and end-to-end training for segmentation tasks. The skip connections solved the problem of losing more precise location information by combining the location information "where" from the encoding part with the semantic information "what" from the decoder through the summation of corresponding pooling and up-sampling layers in two paths. U-Net extended the core idea of FCN by proposing a symmetric encoder–decoder network that differed from FCN in the sense that the skip connections were more present throughout a network with a symmetric structure in the encoder and decoder parts. In addition, instead of summation, U-Net utilized concatenation for the feature maps and further processed the resulting feature maps.

Following the success of U-Net and FCN, in recent years, research [41,47–55] in the medical imaging domain has exceeded the performance limits of segmentation through the adaptation and advancement of these approaches. For instance, Drozdzal et al. [56] explored the idea of creating a deeper FCN by adding a short skip connection to the decoder and encoder paths. Zhou et al. [49] developed more sophisticated skip connections to create features that were more semantically compatible before merging the feature maps from the contacting and expanding paths. Another architecture named UNet++ [57] introduced the idea of using an ensemble of U-Nets with varying depths to address the problem of unknown optimal depth of variations of U-Net and also proposed a new skip connection that enabled the aggregation of features from previous layers as opposed to the traditional way of combining features from the same semantic level.

Hai et al. [58] moved beyond general modification to the U-Net architecture and improved the design while considering the challenging features of the mammography data, such as the diversity of shapes and sizes. To this end, they utilized an Atrous Spatial Pyramid Pooling (ASPP) module in the transition between the encoder and decoder paths. The ASPP block consisted of 1×1conv plus three atrous convolutions [59] with sample rates of 6, 12, and 18; the outputs for these layers were concatenated and fed into a 1×1 conv. FC-DenseNet [60] was selected as the backbone of the method, which is an adaptation of Excellent DenseNet [61] for the segmentation task by constructing a U-Net shape network. Shuyi et al. [62] is another U-Net-based approach based on the idea of utilizing densely connected blocks for mass segmentation. In the encoder, the path is constructed from densely connected CNNs [61]. In the decoder, gated attention [52] modules are used when combining high and low-level features, allowing the model to focus more on the target. Another line of research within the scope of multi-scale studies is [63], in which the generator is an improved version of U-Net. Multi-scale segmentation results were created for three critics with identical structures and different scales in the discriminator. Ravitha et al. [25] developed an approach to use the error of the outputs of intermediate layers (in both encoder and decoder paths) relative to the ground truth labels as a supervision signal to boost the model's performance. In every stage of the encoder and decoder, attention blocks with upsampling were applied to the outputs of the block. The resulting features were linearly combined with the output of the decoder and incorporated into the objective criterion of the network to enhance the robustness of the method.

Sun et al. [21] introduced an asymmetric encoder–decoder network (AU-Net). While in the encoder path, res blocks (three conv layers with a residual connection) were used; in the decode path, basic blocks (including two conv layers) were utilized. The main contribution of the paper was a new upsampling method. In the new attention upsampling (AU) block, high-level features were upsampled through dense and bilinear upsampling. Then, the low-level was combined with the output of dense upsampling through element-wise summation. The resulting feature maps from the previous step were concatenated with the output of bilinear upsampling and were fed into a channel-wise attention module. Finally, the input of the channel-wise attention module was combined with its output by channel-wise multiplication. To address the relatively low performance of U-Net on small-size masses, Xu et al. [22] proposed using a selective receptive field module with two modules—one for generating several receptive fields with different sizes (MRFM) and one for selecting the appropriate size of the receptive field (MSSM) in the down-sampling path of the network. MRFM, in separate paths, applies 3×3 and 5×5 conv on the input image. MSSM takes the two sets of feature maps from the MRFM module, and after combining the features from multi-scale paths in the selection module, the input is processed in three branches with global average pooling and two local context branches using convolutions with different kernel sizes. AU-Net [21] was the baseline method in this study and ARF-Net [22] was used for the comparison.

2.2. Loss Functions for Mass Segmentation

The loss function is one of the essential components in the design of a successful deep-learning approach. For the mass segmentation task, which is designed as pixel-level labeling of the input image, the loss function compares the labels in the ground truth and predicted masks for each pixel to penalize the mismatches. With regard to the mass segmentation task, masses could have various sizes (from less than one percent of the image to considerably larger sizes), shapes, and appearances. One of the greatest challenges for the mass segmentation task is the pixel class imbalance problem. In addition, one extra factor that is specific to mammography is that mammography images have various tissue density categories, and the identification of masses becomes harder as the category number increases. Therefore, in this section, we focus on the losses proposed for binary segmentation for medical images and, specifically, those useful for the aforementioned challenges. Considering the paper's contribution, we categorized the losses into pixel-level

and region-level losses. In this paper, we refer to the loss functions that measure the similarity by only comparing a pair of pixels as pixel-level losses. On the contrary, the second group of losses that incorporate regional information, such as the similarity of the surrounding pixels in the calculation of the loss for a pair of pixels, are referred to as region-level losses.

2.2.1. Pixel-Level Losses

One of the widely used losses in this category is binary cross entropy (BCE) [32], which penalizes mismatch between pixels. As the pixel level computation in BCE makes it vulnerable to bias toward correctly predicting the majority class in the presence of pixel class imbalance, which is specifically common in medical applications, several attempts have been made to ensure the BCE loss is robust to the problem [64]. In this category, focal loss [65] aims to diminish the effect of pixel class imbalance by reducing the impact of easy samples. Additional variants of focal loss can be found in [28,66]. Balanced cross entropy [67,68] controls the impact of classes by using the ratio of pixels in each class to the total as balancing coefficients in BCE terms. Dice loss [69] is robust to the pixel class imbalance problem [70], as it measures correctly classified pixels within true and predicted positive classes. Dice loss is also one of the common losses used for segmentation, and there are several variants of dice loss [29,71,72]. Tversky loss [30] balances the contribution of the false positive and the false negative terms by assigning coefficients to each of the false positive and false negative terms.

2.2.2. Region-Level Losses

In this category, which was initially proposed for image quality assessment, the Structural Similarity Index Measure (SSIM) [73] was utilized in segmentation loss for medical image segmentation [55]. SSIM measures the similarity of images in pixel and region levels. Regional Mutual Information (RMI) [33] and Structural Similarity Loss (SSL) [34] are two examples of region-level losses developed specifically for segmentation. The definition of the region can be refined, as in [74,75], where each region is represented by a cell in a grid, or it is dynamically adjusted, for example through sliding window approaches, to calculate the loss for each of the pixels in the ground truth. It should be noted that both of these losses consider a fixed-size window around each pixel as the region and belong to the latter category.

Inspired by the influence of the structural term in SSIM, the authors of SSL [34] introduced a method for capturing structural similarity by weighting the cross-entropy of every two pixels based on the structural error (error between two regions, which indicates the degree of linear correlation), while ignoring easy pixels and emphasizing pixels with a high error by thresholding the error rate. In the same category, RMI [33] considered a region around a centering pixel as a multi-dimensional point (for a 3×3 region, it will be a 9D point) and then maximized the MI between multidimensional points.

Using two or several losses in a hybrid setting is a common practice in the literature. For instance, all three region-level losses have been tested in a hybrid setting in combination with other losses. In this category, several losses [27,30,31,76] have been proposed to combine beneficial losses for a certain task. As the weighted sum of BCE and dice, combo loss [27] was proposed to control the contribution of false positive and false negative rates by a weighting strategy within the BCE loss term.

3. Materials and Methods

BCE is a widely used loss function for binary segmentation. When applied to mass segmentation in medical imaging, and mammography images specifically, the presence of pixel class imbalance, which is a domain-specific property of the data, can lead to bias. This bias is prone to favoring correctly classifying the major class, as it considers the contribution of both classes as equal while measuring the similarity of the ground-truth and the prediction masks. Therefore, a common practice is to use it in a hybrid setting

alongside other losses that will help to mitigate this problem. For example, dice loss ignores the true negative in the calculation, which makes it immune to favoring the majority class, but opens the door for other issues, such as instability in training. Therefore, combining these losses is beneficial with respect to both losses. In recent years, there has been a surge in hybrid losses for mass segmentation and medical imaging in general due to the proven performance gain they provide. However, they mostly use static weights for the losses and only utilize the information in the ground truth and predicted masks. We argue that hybrid loss could benefit from adaptive weighting, which can be customized for each sample. In addition, we propose using the available information, such as the mass ratio and ACR density, for calculating the loss as a re-weighting signal for the adaptive sample-level prioritizing losses. ACR stands for American College of Radiology (ACR), which developed a standard categorization for density for mammography images containing four categories, from low to high densities. In the following sections, we delve into the details of using mass ratio and ACR density in the proposed loss functions (shown in Figure 1).

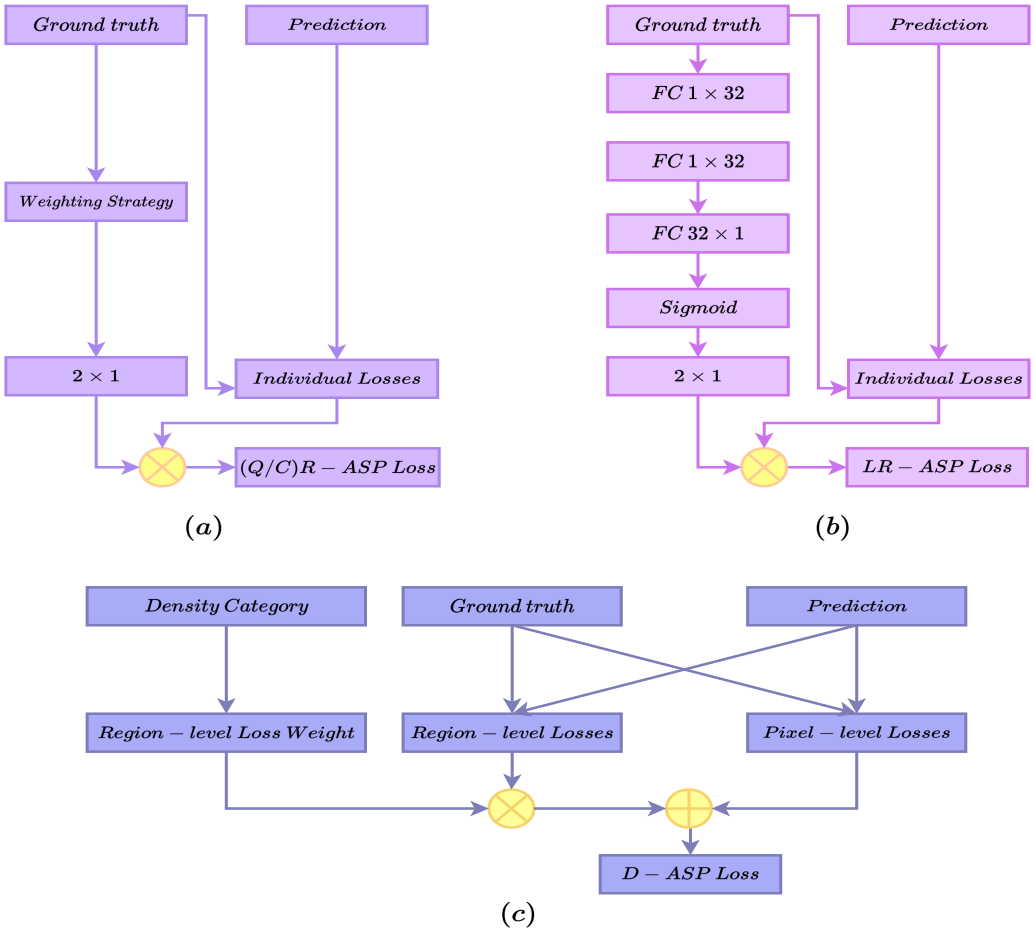

Figure 1. An overview of the proposed ASP losses. (**a**) presents the QR-ASP and CR-ASP losses. (**b**) shows the LR-ASP loss and the sub network for learning the weights. (**c**) presents the D-ASP loss calculation module.

3.1. Ratio as Weighting Signal in the Adaptive Sample-Level Prioritizing Loss

One of the challenges in mass segmentation is that masses can occupy anywhere from less than 1% of the image to a significantly larger area. This means the severity of the pixel class imbalance will vary greatly between the samples. Therefore, the ratio of the mass to the image size can be considered a direct indicator of the severity of the pixel class imbalance problem for individual samples. We propose using this ratio as a sample-level signal to adaptively re-weight losses. One crucial aspect is how to incorporate the adaptive signal into the loss. We propose utilizing the relationship between losses and the severity

of the pixel class imbalance. For instance, BCE is not suitable for small ratios in which the pixel class imbalance is high. On the other hand, dice loss can capture the differences between pixels better for samples with low ratios, in which the pixel class imbalance is high. Therefore, based on this observation, we formulate the Ratio-ASP (R-ASP) loss in a way that BCE has an inverse relationship with the mass ratio of the sample, and the dice has a direct relationship with the ratio, as presented in Equation (1).

$$L^i_{R\text{-}ASP} = (I_{Dice} + (1 - p^i)\gamma)L^i_{Dice} + (I_{BCE} + (p^i)\gamma)L^i_{BCE} \quad (1)$$

Here, $L^i_{R\text{-}ASP}$ is the R-ASP loss for the ith sample. I_{Dice} and I_{BCE}, are initial weights for dice and BCE, respectively. L_{Dice} and L_{BCE} are dice and BCE losses, respectively. p indicates whether the ith sample belongs to the large sample group, and γ is the hypermarapeter controlling the effect of the ratio in R-ASP. It should be noted that using the raw ratio will make the training unstable, as weighting will be sensitive to even a very small amount of change. Hence, we propose grouping the samples into two categories, small and large, based on the ratio of the mass. In this regard, the dividing value and strategy becomes important. In this section, we introduce three different ways to divide the samples into two categories (calculating p^i).

3.1.1. Quantile-Based R-ASP Loss

An intuitive way to divide the samples according to the ratios is to divide them according to the quantile to which they belong based on the value of the median, as shown in Equation (2), which we name Quantile-based R-ASP (QR-ASP).

$$p^i_Q = \begin{cases} 0 & r^i \in Q\_1 \\ 1 & \text{Otherwise} \end{cases} \quad (2)$$

Here, r^i is the ratio of the ith sample and Q_1 is the first quantile. In the ablation study section, we also tested this variation with the mean value of the ratios for all of the samples as a dividing factor; for future reference, we used the term Value-based ASP (VR-ASP) for these experiments.

3.1.2. Cluster-Based R-ASP Loss

The distribution of the data is skewed for both of the datasets; hence, we speculated that QR-ASP might not have been the best solution in such scenarios, which were sampled from real-world data. Therefore, in this variation, we proposed dividing the samples based on the cluster to which each sample belonged, as in Equation (3), which we termed Cluster-based R-ASP (CR-ASP). Figure 1a presents the QR/CR-ASP variants.

$$p^i_C = \begin{cases} 0 & r^i \in C_s \\ 1 & \text{Otherwise} \end{cases} \quad (3)$$

Here, C_s is the center of the cluster for the small ratio group. The advantage of CR-ASP was that, in this case, the division of the samples was a better representative of the distribution of data in the case of skewed data compared with using the QR-ASP method.

3.1.3. Learning-Based R-ASP Loss

Finally, in the last version of R-ASP, we proposed learning the weights for the hybrid loss. To this end, a parameterized subnetwork was used who used the inputs of ground truth, and its outputs were the weights for the BCE and dice losses. The subnetwork is presented in Figure 1b and the loss is presented in Equation (4), which we named learning-based R-ASP (LR-ASP). We speculate that learning the weights during the training might provide the method with an automated way to capture the differences between samples.

$$L^i_{LR\text{-}ASP} = f(y, w)_1 L^i_{Dice} + f(y, w)_2 L^i_{BCE} \quad (4)$$

Here $f(y, w)_1$ and $f(y, w)_2$ are the outputs of the module that learn the relation between the samples and loss terms, respectively. w presents the weights of the LR-ASP module.

3.2. ACR Density as Weighting Signal in the Hybrid Adaptive Sample-Level Loss

While the ratio of mass assists the loss in handling the class imbalance problem, there is another problem specific to mammography images that makes the segmentation task challenging, which is the ACR density category of the breast. In samples with more dense tissues, the identification of masses is more demanding, and, in some cases, normal tissues might hide the masses. In most cases, the proposed methods and losses do not take ACR density into consideration for mass segmentation, while it is available for each examination (using the ACR or other methods for labeling the density). In this section, we proposed a new hybrid loss function that included the region-level and pixel-level losses. The pixel-level losses were sufficient for less dense samples, while region-level losses were more appropriate for samples with higher-density categories. This was because considering the neighboring pixels in loss calculation helped with better segmentation for more dense samples, in which the segmentation was more challenging. In order to signal the loss to prioritize one of the loss terms according to the density category of the sample, we proposed using ACR density as an adaptive signal for the hybrid loss function (including pixel-level and region-level losses).

3.2.1. Pixel-Level Loss Term

As mentioned before, we referred to losses that considered the pixels as independent entities as pixel-level losses. From this category, we used a weighted sum of BCE and dice loss, which is a commonly used hybrid pixel-level loss. The pixel-level loss was prioritized more for the samples that had a lower ACR density and provided the benefit of both BCE (Equation (5)) and dice (Equation (6)) losses to some extent. The pixel-level term is presented in Equation (7).

$$L_{BCE} = -\Big(y log(\hat{y}) + (1-y) log(1-\hat{y})\Big) \tag{5}$$

$$L_{Dice} = 1 - \frac{\sum_{j=1}^{H\times W} \hat{y}_j y_j + \epsilon}{\sum_{j=1}^{H\times W} \hat{y}_j + \sum_{j=1}^{H\times W} y_j + \epsilon} \tag{6}$$

$$L_{HP} = \alpha L_{Dice} + \beta L_{BCE} \tag{7}$$

Here, y and $\hat{y}$ are the ground truth and the predicted segmentation, respectively. α and β are the weighting parameters in the hybrid loss denoted as L_{HP} in Equation (7). In our experiments, both α and β were set to one. W and H are the width and height of the images, respectively.

3.2.2. Region-Level Loss Term

Region-level loss in this paper refers to the category of losses that incorporate the surrounding pixels in the calculation of the loss for a centering pixel. We selected SSIM and RMI losses that, based on our experiments, performed well in a combined setting. It should be noted that ACR density only had a connection with the region-level losses; therefore, it was only used for prioritizing the region-level term. One of the region-level losses was RMI in Equation (8), which measured the similarity of multi-dimensional points generated by comparing a centering pixel and its surrounding pixels (in the ground truth and predicted masks), which allowed the method to achieve consistency between the predicted and true mask beyond what using only the pixel-level comparison could measure.

$$L_{RMI}(Y_m; \hat{Y_m}) = \int_S \int_{\hat{S}} f(y, \hat{y}) \log\left(\frac{f(y, \hat{y})}{f(y) f(\hat{y})}\right) dy\, d\hat{y} \tag{8}$$

Here, Y_m and $\hat{Y_m}$ are the multi-dimensional points that are generated from a window with a fixed size around a centering pixel capturing the dependency of the neighboring pixels. The subscript m refers to the 'multi-dimensional' term. S and $\hat{S}$ are the support sets for the ground truth and prediction masks, respectively. $f(y)$ and $f(\hat{y})$ represent the probability density functions of the ground truth and prediction masks, respectively. $f(y, \hat{y})$ computes the joint PDF. The implementation details of the RMI loss are available in [33] and the code is publicly available.

The second region-level loss used in this paper was SSIM-based loss, as shown in Equation (9).

$$L_{SSIM}(Y_p; \hat{Y}_p) = 1 - \frac{(2\mu_{y_p}\mu_{\hat{y}_p} + C_1)(2\sigma_{y\hat{y}} + C_2)}{(\mu_{y_p}^2\mu_{\hat{y}_p}^2 + C_1)(\sigma_y^2 + \sigma_{\hat{y}}^2 + C_2)} \tag{9}$$

Here, Y_p and $\hat{Y}_p$ represent patches in the ground truth and prediction masks, respectively, in which subscript p refers to the 'patch' term. 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 C_1 and C_2) are available in the original paper [73]. Finally, the hybrid region-level loss is presented in Equation (10).

$$L_{HR} = \eta L_{RMI} + \tau L_{SSIM} \tag{10}$$

In the hybrid region-level loss L_{HR} (Equation (10)), L_{RMI} and L_{SSIM} are the RMI and SSIM losses, respectively. The hyperparameters η and τ represent the weighting coefficients. Figure 1c and Equation (11) present the density-based ASP (D-ASP) loss.

$$L_{D\text{-}ASP}^i = d^i\theta^T L_{HR}^i + L_{HP}^i \tag{11}$$

$L_{D\text{-}ASP}^i$ is the final D-ASP loss for the ith sample. θ is the prioritizing vector of the weights assigned to each density category, and d^i denotes a one-hot encoding of the density category of the ith sample. $d^i\theta^T$ will re-weight the region-level loss term, which determines the importance of the region-level term according to the density category of the ith sample.

3.3. Adaptive Sample-Level Focal Loss

Up to this point, we have investigated the idea of using size and ACR density as the re-weighting signal in hybrid losses. In this section, we aim to explore the possibility of using the ratio category as a sample-level value to control the impact of pixels in the samples in a single loss. To this end, we selected focal loss, which was appropriate for this purpose as it was based on BCE and was developed for the same purpose. In the focal loss presented in Equation (12), υ is the focusing parameter, which reduces the impact of loss for easy-to-classify pixels. Assuming that the segmentation of the larger masses was easier as they are more salient compared with the smaller masses, we proposed increasing υ for the large category and decreased it for the small category. This enabled the loss to adaptively change the hyper-parameter that was considered to be fixed for all samples in the in the original focal loss.

$$L_F(\hat{Y}_t) = -(1 - \hat{Y}_t)^{\upsilon} \cdot \log(\hat{Y}_t) \tag{12}$$

Here, υ is the focusing parameter and $\hat{Y}_t$ presents the model's estimated probability for the mass. In the modified focal loss, there will be two values for focusing parameters, corresponding to the large and small groups instead of a fixed value, as presented in Equation (13). This, in turn, allows for the sample-level adaptation of focal loss according to the property of the sample.

$$\upsilon_{ASP} = \begin{cases} \upsilon_{c_1} & r^i \in C_1 \\ \upsilon_{c_2} & r^i \in C_2 \end{cases} \tag{13}$$

Here, c_1 and c_2 are the two categories for the samples and r^i is the ratio of the ith sample. We used the name Single (loss) R-ASP (SR-ASP) loss. v_{ASP} presents the adaptive v. The same idea could be used for ACR density, as the density of the sample was related to the hardness of the segmentation. This variant was called Single (loss) D-ASP (SD-ASP). It should be noted that the ratio and density were not directly connected to each pixel, but to each image. Therefore, the ASP version of the focal loss indicated that the effect of an easy pixel in an easy sample should be lower and vice versa.

3.4. Evaluation Metrics

Dice Similarity Coefficient, (DSC, Equation (14)), Relative Area Difference, (ΔA, Equation (15)), Sensitivity (Equation (16)), and Accuracy (Equation (17)) were selected as the evaluation metrics in all of the experiments due to the complementary information they provided. As masses occupied a trivial portion of the images, using accuracy solely would not be an informative mean to evaluate the method. Hence, using an additional metric such as sensitivity that allowed us to measure the false negative rate, which is important in medical applications, is vital. DSC measures the ratio of the correctly predicted positive pixels over the number of positive areas in both the ground truth and the prediction mask, considering the false positive rate in the calculations alongside false negatives. Therefore, we also included DSC in the evaluation metrics. Finally, the relative area difference, which measures the difference in the sizes of the actual and predicted masses compared to the size of true mass (smaller value for ΔA, indicates closer sizes for masses and is better), was considered in the experiments. In all of the tables, the best results are highlighted in bold font. In the figures for the examples of the segmentation results, the green and blue borders present the ground truth and the prediction of the methods, respectively.

$$DSC = \frac{2TP}{2TP + FP + FN} \tag{14}$$

$$\Delta A = \frac{|(TP + FP) - (TP + FN)|}{TP + FN} \tag{15}$$

$$Sensetivity = \frac{TP}{TP + FN} \tag{16}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FN + FP} \tag{17}$$

Here, TP, TN, FP, and FN represent true positive, true negative, false positive, and false negative rates, respectively.

3.5. Datasets and Experimental Setting

INbreast and CBIS-DDSM were used in this paper to test the proposed method. Normalization and resizing to 256×256 were performed on both datasets without extra data augmentation or image enhancement. For the baseline approach, the batch size was set to 4; the learning rate was set to 0.0001 in all experiments. For INbreast, from the 410 images (for 115 cases), 107 images containing masses were used in this study. As the dataset was small, following a common practice for using INbreast, we used a 5-fold cross-validation. The dataset was randomly divided into training (80%), validation (10%), and test (10%) sets. CBIS-DDSM had a total of 1944 cases, in which 1591 images contained masses that were utilized in our experiments. We used the official train (1231) and test (360) split of the dataset with 10% of the training data as a validation set, which was randomly selected. The CBIS-DDSM images had some artifacts; therefore, the artifacts were removed in the preprocessing step.

3.6. Comparison of Dataset Characteristics

INbreast and CBIS-DDSM are different in many respects, ranging from the dataset size to the accuracy of the segmentation. However, regarding the mass size that was used in R-ASP, it was important to analyze the differences in the distribution of the sizes in the two datasets. To this end, the distributions of mass ratios in the datasets are depicted in Figure 2. CBIS-DDSM (Figure 2a) has a smaller range compared with the INbreast dataset (Figure 2b). The smaller length of the interquartile for the CBIS-DDSM dataset indicated less diversity in the ratio for the majority of the masses. In addition, a closer mean and median in the CBIS-DDSM dataset could be a sign of less skewed data for ratio. When comparing the performance of R-ASP, considering these differences could be helpful; regarding D-ASP, the difference in the distribution of the density categories is relevant information that is provided in Figure 3. Generally, for both of the datasets, most of the samples had lower densities.

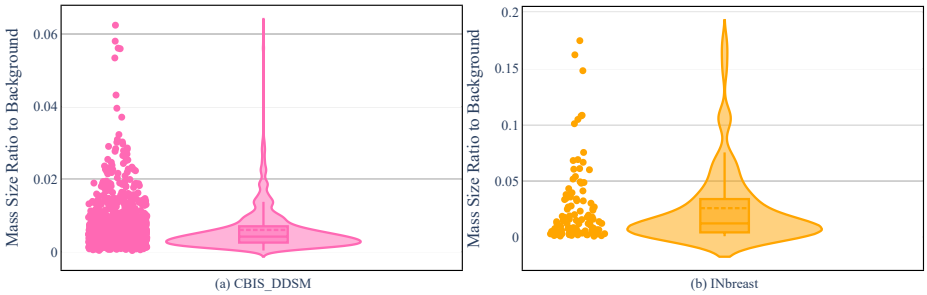

Figure 2. Violin plots for the CBIS-DDSM (**a**) and INbreast (**b**) based on the mass ratio.

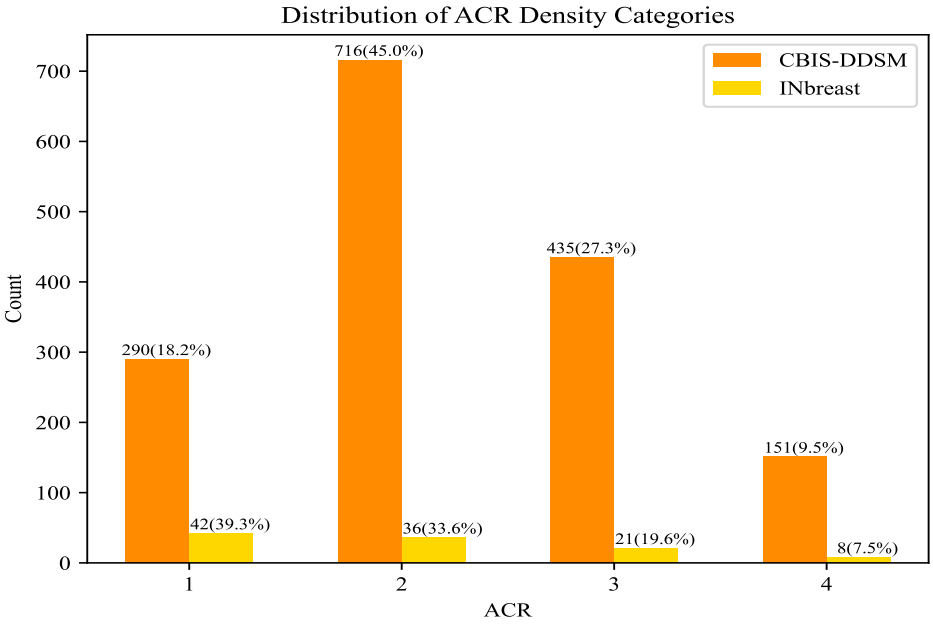

Figure 3. Comparison of INbreast and CBIS-DDSM based on the distribution of the density categories. ACR density increases according to the category number.

4. Experimental Results

4.1. Ablation Study

For CR-ASP, the number of centers was a hyperparameter. We tested CR-ASP with two and three clusters to select the right value. As shown in Figure 4, using two clusters showed a better performance in comparison with three clusters.

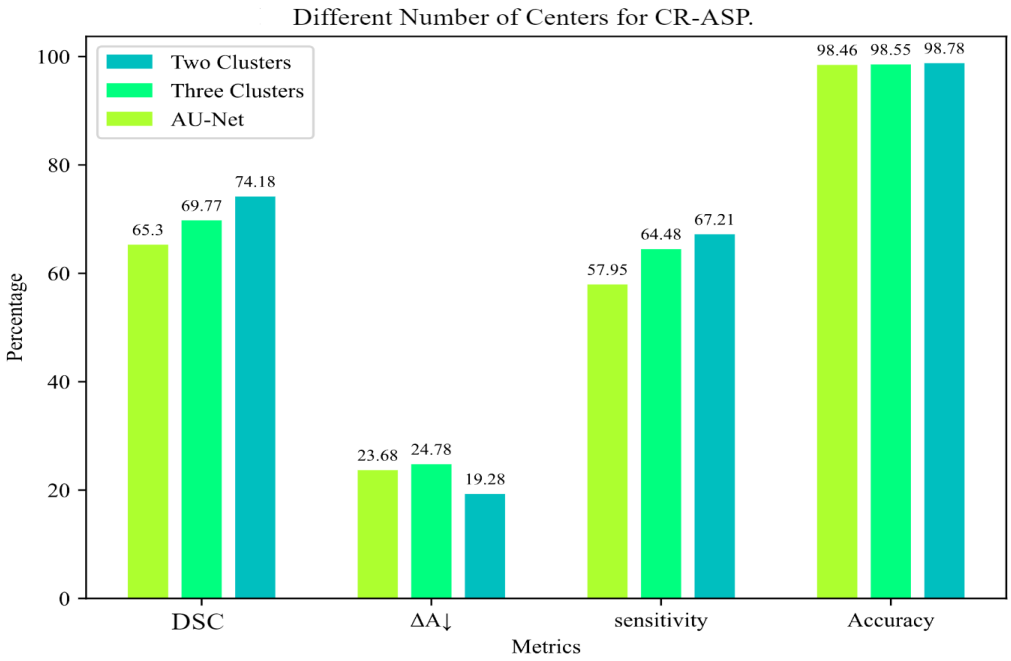

Figure 4. Comparison of different numbers of clusters for CR-ASP loss for INbreast dataset.

Regarding VR-ASP, SR-ASP, and SD-ASP, we conducted experiments on the INbreast dataset, and the results are presented in Table 1. As shown in the table, VR-ASP, with a mean value of 0.025, provided a modest improvement in the baseline method. Regarding VR-ASP, in which the mean value of the ratio was utilized for grouping the samples while improving on the baseline method (shown by the name hybrid in Table 1), VR-ASP achieved the lowest improvement in DSC and accuracy in the R-ASP category, but the sensitivity and ΔA were better than both LR-ASP and QR-ASP. Regarding SD-ASP and SR-ASP, we conducted an experiment to identify the best υ value for the focal loss, which was 0.5 for our experiments. SR-ASP and SD-ASP provided a considerable boost in the performance of the focal loss in a single loss setting, as shown in Table 1. SR-ASP surpassed focal loss in all the metrics and performed better than SD-ASP in terms of DSC and accuracy. It should be noted that while DSC was better for SR-ASP compared with SD-ASP, the sensitivities were comparable. This might indicate a better reduction in false negative rate for the SD-ASP, which was aligned with the experimental results in the following sections.

For the SR-ASP, the value of υ was set to 0.25 for samples belonging to the small group, and 0.5 for the large group. For SD-ASP, the values were [0.2, 0.25, 0.3, 0.35] for categories from 1 to 4, respectively. For VR-ASP, the values for γ in Equation (2) were 0.25 and 0.25 and the I_{Dice} and I_{BCE} were considered one.

Table 1. Results for experiments for SR-ASP, SD-ASP, and VR-ASP on INbreast. In each subsection of the table, the best results for each metric have been shown in bold.

Losses	DSC	ΔA	Sensitivity	Accuracy
Focal	54.67	46.8	42.49	98.19
SR-ASP	**68.83**	26.63	**62.90**	**98.55**
SD-ASP	66.79	**16.41**	62.68	98.47
Hybrid	65.32	23.68	57.95	**98.46**
VR-ASP	**66.94**	**20.81**	**64.77**	98.43

4.2. Comparison with State-of-the-Art Method

In order to examine the strengths and weaknesses of the proposed R-ASP and D-ASP losses, they were tested on INbreast and CBIS-DDSM datasets. AU-Net and ARF-Net

were selected for comparison. AU-Net was the baseline method; thus, it was essential to compare the performance of ASP variations with it. As AU-Net used the hybrid loss function, including dice and BCE, comparison with AU-Net provided information about the impact of ASP losses. ARF-Net is another method in which different scales were incorporated into the design of the network; this was related to our study as it incorporated the idea of designing the network in a way that leveraged different sizes. For AU-Net, the official implementation was publicly available, and the training setting was closely followed by the description in the paper. We implemented ARF-Net to the best of our understanding of the ARF-Net paper. For the region-level term in D-ASP, publicly available implementations of RMI and SSIM were utilized. No pre-training or data augmentation were used in any of the experiments. For R-ASP the hyperparameters I_{Dice} and I_{BCE} were set to 0.125 and 0.25 for INbreast and CBIS-DDSM, respectively. For γ, a value between [0.25, 0.35] was used with different values for the dice and BCE terms. For the cluster-based strategy, the number of centers was a hyperparameter, for which 2 was selected. Regarding D-ASP, the coefficients were $\theta = [0.5, 0.5, 0.85, 0.95]$ and $\theta = [0.25, 0.25, 0.85, 0.95]$ for the INbreast and CSIB-DDSM, respectively. η, τ, β were set to 1; α was set to 2 and 2.5 for INbreast and CBIS-DDSM, respectively. All of the hyperparameters were selected through experimental evaluation.

4.2.1. Experimental Results for R-ASP

Tables 2 and 3 presented the results of the experiments for R-ASP using the INbreast and CBIS-DDSM datasets, respectively. For the INbreast dataset, for the best-performing variation, CR-ASP, the improvements were as follows: DSC: +8.86%, ΔA: −4.4%, Sensitivity: +9.26%, and Accuracy: +0.32%. Considering Figure 2, due to the distribution of the sizes in the dataset, as expected, the CR-ASP version outperformed QR-ASP. While QR-ASP improved on the baseline generally, we speculate that its subpar performance compared with the two other R-ASP variations could be due to the fact that the groupings of the samples to small and large categories using QR-ASP were not suitable considering the distribution of the sizes in the dataset. Regarding LR-ASP, while it did not boost the performance to the same level as CR-ASP, it surpassed AU-Net, ARF-Net, and QR-ASP in all of the metrics. The achieved level of performance for LR-ASP compared with the baseline indicated that the learned weights for the loss terms were useful for prioritizing the loss terms. In general, CR-ASP seemed to provide a better connection between the samples and the loss terms.

Table 2. Comparison of R-ASP with state-of-the-art approaches for INbreast. The best results for each metric have been shown in bold.

Method	DSC	$\Delta A \downarrow$	Sensitivity	Accuracy
ARF-Net	70.05	30.37	59.59	98.71
AU-Net	65.32	23.68	57.95	98.46
QR-ASP	68.03	25.04	63.12	98.54
LR-ASP	71.92	22.31	64.56	98.71
CR-ASP	**74.18**	**19.28**	**67.21**	**98.78**

Table 3. Comparison of R-ASP with state-of-the-art approaches for CBIS-DDSM. The best results for each metric have been shown in bold.

Method	DSC	$\Delta A \downarrow$	Sensitivity	Accuracy
ARF-Net	48.82	11.47	47.27	99.43
AU-Net	49.05	09.94	51.49	99.38
QR-ASP	**51.48**	**02.05**	**52.00**	99.43
LR-ASP	51.33	23.17	45.38	**99.50**
CR-ASP	51.04	04.47	49.90	99.45

Regarding CBIS-DDSM, while the best results for all of the metrics were achieved by variations of R-ASP, QR-ASP was the best-performing variation, as shown in Table 3. This showed that while being generally effective, the degree of effectiveness was related to the appropriate grouping method, given the distribution of data. Figure 5 presents a few samples of the performance of different approaches for small and large size masses for both of the datasets. The best-performing R-ASP and D-ASP versions provided improvements over the baseline method for INbreast and CBIS-DDSM.

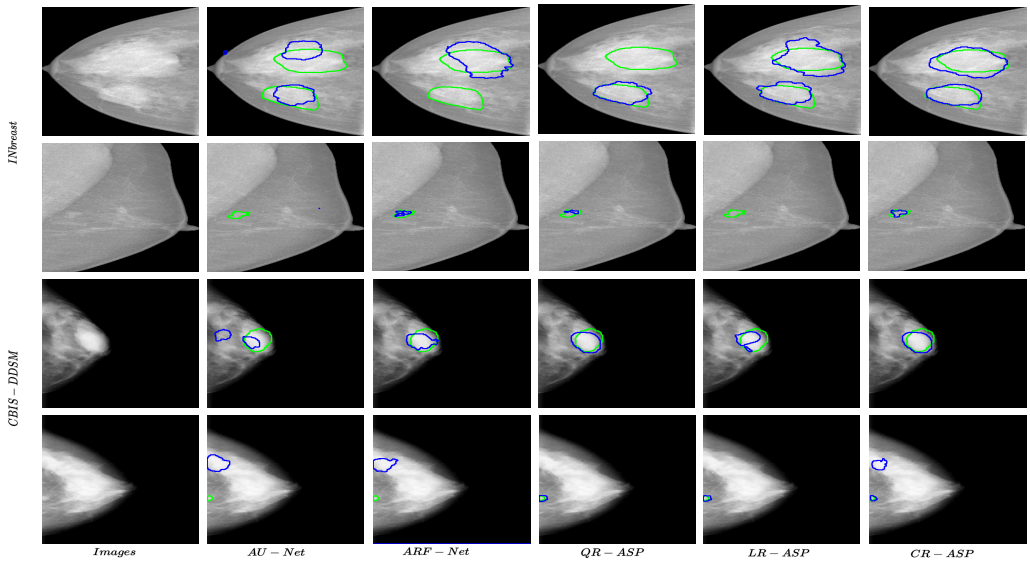

Figure 5. Results for R-ASP variants, AU-Net, and ARF-Net approaches for both datasets. The green color shows the border of ground truth and the blue color presents the prediction.

4.2.2. Experimental Results for D-ASP

The results for D-ASP on INbreast and CBIS-DDSM are summarized in Tables 4 and 5, respectively. D-ASP outperformed other methods using traditional pixel-level hybrid loss functions. Notably, the improvements over the baseline hybrid losses were as follows: DSC: $+9.27\%$, ΔA: -12.77%, Sensitivity: $+20.21\%$, Accuracy: +0.19 %. Compared with R-ASP, D-ASP performed better in terms of DSC, ΔA, and sensitivity. For the CBIS-DDSM dataset, D-ASP also improved on the baseline method and ARF-Net (Table 5). However, the results were comparable to R-ASP. One observation was that D-ASP outperformed R-ASP in sensitivity on both datasets. For instance, when the DSCs were comparable, the sensitivity was considerably higher for D-ASP. The reason could be that the reduction in false positives was stronger in R-ASP, while in D-ASP, false negatives were more diminished. It should be noted that while both R-ASP and D-ASP shared the main idea of prioritizing the loss terms, they differed in the information that was used for the re-weighting signal and the losses used. In addition, for R-ASP, the re-weighting depended on the samples' ratios and the interpretation of the ratios in a group of samples (training set). On the other hand, for D-ASP, the re-weighting signal depended on the information of the sample itself. Therefore, the difference in performance could be attributed to all of these factors.

Figure 6 presents one example from each ACR category for both datasets. One observation in all the samples was that the borders of D-ASP's segmentations matched the borders of the true mass better compared with all of the other methods and R-ASP. This could be because incorporating the region-level losses enabled the loss to include the context in the calculation and eliminated false predictions.

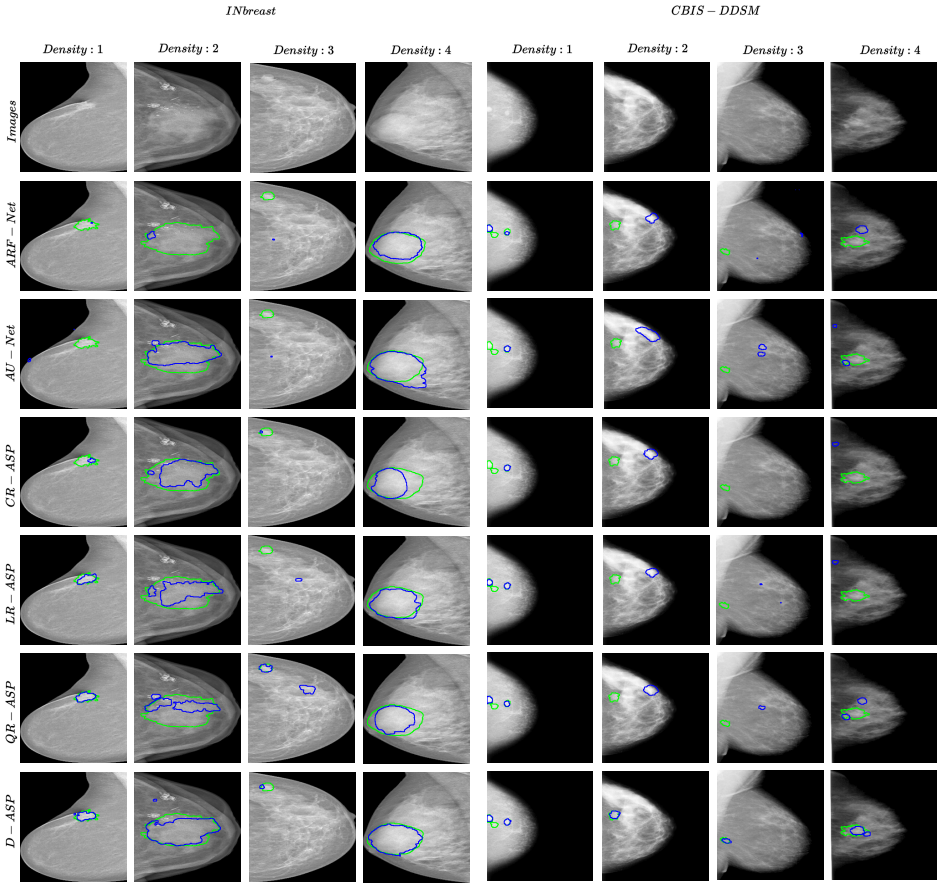

Figure 6. Examples of segmentation results for ARF-Net, AU-Net, R-ASP, and D-ASP for both datasets. One sample from each ACR density category is included. The green color shows the border of ground truth and the blue color presents the prediction.

Table 4. Results for D-ASP, R-ASP and state-of-the-art approaches for INbreast. The best results for each metric have been shown in bold.

Method	DSC	$\Delta A \downarrow$	Sensitivity	Accuracy
ARF-Net	70.05	30.37	59.59	98.71
AU-Net (baseline)	65.32	23.68	57.95	98.46
QR-ASP	68.03	25.04	63.12	98.54
LR-ASP	71.92	22.31	64.56	98.71
CR-ASP	74.18	19.28	67.21	**98.78**
D-ASP	**74.59**	**10.91**	**78.16**	98.65

Table 5. Results for D-ASP, R-ASP and state-of-the-art approaches for CBIS-DDSM. The best results for each metric have been shown in bold.

Method	DSC	$\Delta A \downarrow$	Sensitivity	Accuracy
ARF-Net	48.82	11.47	47.27	99.43
AU-Net (baseline)	49.05	09.94	51.49	99.38
QR-ASP	**51.48**	**02.05**	52.00	99.43
LR-ASP	51.33	23.17	45.38	**99.50**
CR-ASP	51.04	04.47	49.90	99.45
D-ASP	50.64	05.96	**52.15**	99.41

5. Discussion

While ASP losses enhanced the performance of the baseline method, there were a few challenges that might have limited their impact. For instance, R-ASP variations depended on the grouping strategy; therefore, the performance could be improved by selecting the appropriate grouping strategies. Regarding D-ASP, one observation is that the ACR density category in some samples might not have been aligned with the expectation of visual differences, which was important for region-level losses. Hence, an inaccurate assessment of the density or any mismatch might have limited the performance of D-ASP. For both the ratio and density-based ASPs, the accuracy of the segmentation mask was important. This was a limitation regarding the CBIS-DDSM dataset, in which the segmentation masks were less accurate compared with INbreast. Regarding the computational time, the proposed method introduced a trivial additional time to the training, as the ratio and density had already been calculated for the training data and only re-weighting was performed during the training time.

6. Conclusions

In this paper, we propose using additional information in each sample as a re-weighting signal for prioritizing one of the loss types, which is more useful for the sample in an adaptive manner. We introduce two variations for the proposed loss. In one variation, called R-ASP, the ratio of the size of the mass to the image size is used for prioritizing dice loss over BCE, and vice versa. This version is based on the assumption that dice loss could be more useful for handling pixel class imbalance, and BCE could be more useful for samples with less severe pixel class imbalance. In the second variation, the density of each sample is utilized as a re-weighting signal in a hybrid loss with pixel-level and region-level loss terms. The rationale behind this variation is the fact that segmentation in samples with a higher ACR density is more challenging; therefore, region-level losses that utilize the information in the neighboring pixels while calculating the loss value for a central pixel can be beneficial for samples with a high ACR density. Both R-ASP and D-ASP are proposed for hybrid losses. In order to use the sample-level adaptation of ASP, we also introduce a variation in focal loss that uses the sample-level information (ratio and density) for the selection of the focusing parameter.

According to the results on both datasets for APS losses, the idea of using sample-level prioritizing loss in addition to including the region-level loss is a promising approach for improvement of the training, and it results in a boost in performance. Using ASP for focal loss also results in a promising performance enhancement. Both R-ASP and D-ASP loss could be generalized for other tasks in medical imaging.

Author Contributions: Conceptualization, P.A., M.N. (Mircea Nicolescu), M.N. (Monica Nicolescu) and G.B.; methodology, P.A.; software, P.A.; validation, P.A.; formal analysis, P.A.; investigation, P.A.; resources, P.A.; data curation, P.A.; writing—original draft preparation, P.A.; writing—review and editing, P.A., M.N. (Mircea Nicolescu), M.N. (Monica Nicolescu) and G.B.; visualization, P.A., M.N. (Mircea Nicolescu), M.N. (Monica Nicolescu) and G.B.; supervision, M.N. (Mircea Nicolescu), M.N. (Monica Nicolescu) and G.B.; project administration, P.A., M.N. (Mircea Nicolescu), M.N. (Monica Nicolescu) and G.B.; funding acquisition, M.N. (Mircea Nicolescu), M.N. (Monica Nicolescu) and G.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: INbreat and CBIS-DDSM datasets are publicly available at https://www.kaggle.com/datasets/tommyngx/inbreast2012 (accessed on 1 January 2023) and https://wiki.cancerimagingarchive.net/pages/viewpage.action?pageId=22516629 (accessed on 1 January 2023), respectively.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Henley, S.J.; Ward, E.M.; Scott, S.; Ma, J.; Anderson, R.N.; Firth, A.U.; Thomas, C.C.; Islami, F.; Weir, H.K.; Lewis, D.R.; et al. Annual report to the nation on the status of cancer, part I: National cancer statistics. *Cancer* **2020**, *126*, 2225–2249. [CrossRef] [PubMed]
2. Niu, J.; Li, H.; Zhang, C.; Li, D. Multi-scale attention-based convolutional neural network for classification of breast masses in mammograms. *Med. Phys.* **2021**, *48*, 3878–3892. [CrossRef]
3. Sinthia, P.; Malathi, M. An effective two way classification of breast cancer images: A detailed review. *Asian Pac. J. Cancer Prev.* **2018**, *19*, 3335.
4. Rajaguru, H.; SR, S.C. Analysis of decision tree and k-nearest neighbor algorithm in the classification of breast cancer. *Asian Pac. J. Cancer Prev.* **2019**, *20*, 3777. [CrossRef] [PubMed]
5. Kulshreshtha, D.; Singh, V.P.; Shrivastava, A.; Chaudhary, A.; Srivastava, R. Content-based mammogram retrieval using k-means clustering and local binary pattern. In Proceedings of the 2017 2nd International Conference on Image, Vision and Computing (ICIVC), Chengdu, China, 2–4 June 2017 ; pp. 634–638.
6. Giger, M.L.; Karssemeijer, N.; Schnabel, J.A. Schnabel. Breast image analysis for risk assessment, detection, diagnosis, and treatment of cancer. *Annu. Rev. Biomed. Eng.* **2013**, *15*, 327–357. [CrossRef] [PubMed]
7. Li, L.; Clark, R.A.; Thomas, J.A. Computer-aided diagnosis of masses with full-field digital mammography. *Acad. Radiol.* **2002**, *9*, 4–12. [CrossRef] [PubMed]
8. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **2012**, *25*, 84–90. [CrossRef]
9. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
10. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1–9.
11. Szegedy, C.; Ioffe, S.; Vanhoucke, V.; Alemi, A. Inception-v4, inception-resnet and the impact of residual connections on learning. In Proceedings of the AAAI Conference on Artificial Intelligence, San Francisco, CA, USA, 4 February 2017; Volume 31.
12. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference, Munich, Germany, 5–9 October 2015; pp. 234–241.
13. Singh, V.K.; Rashwan, H.A.; Romani, S.; Akram, F.; Pandey, N.; Sarker, M.K.; Saleh, A.; Arenas, M.; Arquez, M.; Puig, D.; et al. Breast tumor segmentation and shape classification in mammograms using generative adversarial and convolutional neural network. *Expert Syst. Appl.* **2019**, *139*, 112855. [CrossRef]
14. Yan, Y.; Conze, P.H.; Quellec, G.; Lamard, M.; Cochener, B.; Coatrieux, G. Two-stage multi-scale mass segmentation from full mammograms. In Proceedings of the 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI), Nice, France, 13–16 April 2021; pp. 1628–1631.
15. Yan, Y.; Conze, P.-H.; Decenciere, E.; Lamard, M.; Quellec, G.; Cochener, B.; Coatrieux, G. Cascaded multi-scale convolutional encoder-decoders for breast mass segmentation in high-resolution mammograms. In Proceedings of the 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Berlin, Germany, 23–27 July 2019; pp. 6738–6741.
16. 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.* **2012**, *19*, 236–248. [CrossRef]
17. Liu, Y.; Zhou, C.; Zhang, F.; Zhang, Q.; Wang, S.; Zhou, J.; Sheng, F.; Wang, X.; Liu, W.; Wang, Y.; et al. Compare and contrast: Detecting mammographic soft-tissue lesions with C2-Net. *Med. Image Anal.* **2021**, *71*, 101999. [CrossRef] [PubMed]
18. Liu, Y.; Zhang, F.; Chen, C.; Wang, S.; Wang, Y.; Yu, Y. Act like a radiologist: Towards reliable multi-view correspondence reasoning for mammogram mass detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *44*, 5947–5961. [CrossRef] [PubMed]
19. Min, H.; Wilson, D.; Huang, Y.; Liu, S.; Crozier, S.; Bradley, A.P.; Chandra, S.S. Fully automatic computer-aided mass detection and segmentation via pseudo-color mammograms and mask R-CNN. In Proceedings of the 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), Iowa City, IA, USA, 3–7 April 2020; pp. 1111–1115.
20. Soltani, H.; Amroune, M.; Bendib, I.; Haouam, M.Y. Breast cancer lesion detection and segmentation based on mask R-CNN. In Proceedings of the 2021 International Conference on Recent Advances in Mathematics and Informatics (ICRAMI), Ebessa, Algeria, 21–22 September 2021; pp. 1–6.
21. 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.* **2020**, *65*, 055005. [CrossRef]
22. 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* **2022**, *71*, 103178. [CrossRef]
23. Xu, S.; Adeli, E.; Cheng, J.; Xiang, L.; Li, Y.; Lee, S.; Shen, D. Mammographic mass segmentation using multichannel and multiscale fully convolutional networks. *Int. J. Imaging Syst. Technol.* **2020**, *30*, 1095–1107. [CrossRef]
24. Mohamed, A.A.; Berg, W.A.; Peng, H.; Luo, Y.; Jankowitz, R.C.; Wu, S. A deep learning method for classifying mammographic breast density categories. *Med. Phys.* **2018**, *45*, 314–321. [CrossRef]
25. Rajalakshmi, N.; Ravitha, R.; Vidhyapriya, N.; Elango; Ramesh, N. Deeply supervised u-net for mass segmentation in digital mammograms. *Int. J. Imaging Syst. Technol.* **2021**, *31*, 59–71.

26. Zeng, Y.; Chen, X.; Zhang, Y.; Bai, L.; Han, J. Dense-U-Net: Densely connected convolutional network for semantic segmentation with a small number of samples. In Proceedings of the Tenth International Conference on Graphics and Image Processing (ICGIP 2018), Chengdu, China, 12–14 December 2019; Volume 11069, pp. 665–670.
27. 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. *Comput. Med. Imaging Graph.* **2019**, *75*, 24–33. [CrossRef]
28. Abraham, N.; Khan, N.M. A novel focal tversky loss function with improved attention u-net for lesion segmentation. In Proceedings of the 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), Venice, Italy, 8–11 April 2019; pp. 683–687.
29. Sudre, C.H.; Li, W.; Vercauteren, T.; Ourselin, S.; Jorge Cardoso, M. Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. In Proceedings of the Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: Third International Workshop, DLMIA 2017, and 7th International Workshop, ML-CDS, Québec City, QC, Canada, 14 September 2017.
30. Salehi, S.S.M.; Erdogmus, D.; Gholipour, A. Tversky loss function for image segmentation using 3D fully convolutional deep networks. In Proceedings of the International Workshop on Machine Learning in Medical Imaging, Québec City, QC, Canada, 10 September 2017; pp. 379–387.
31. Yeung, M.; Sala, E.; Schönlieb, C.-B.; Rundo, L. Unified focal loss: Generalising dice and cross entropy-based losses to handle class imbalanced medical image segmentation. *Comput. Med. Imaging Graph.* **2022**, *95*, 102026. [CrossRef]
32. Ma, Y.-d.; Liu, Q.; Qian, Z.-b. Automated image segmentation using improved PCNN model based on cross-entropy. In Proceedings of the 2004 International Symposium on Intelligent Multimedia, Video and Speech Processing, Hong Kong, China, 20–22 October 2004; pp. 743–746.
33. Zhao, S.; Wang, Y.; Yang, Z.; Cai, D. Region mutual information loss for semantic segmentation. *Adv. Neural Inf. Process. Syst.* **2019**, *32*. [CrossRef]
34. Zhao, S.; Wu, B.; Chu, W.; Hu, Y.; Cai, D. Correlation maximized structural similarity loss for semantic segmentation. *arXiv* **2019**, arXiv:1910.08711.
35. 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* **2017**, *4*, 170177. [CrossRef] [PubMed]
36. Soulami, K.B.; Kaabouch, N.; Saidi, M.N.; Tamtaoui, A. Breast cancer: One-stage automated detection, segmentation, and classification of digital mammograms using UNet model based-semantic segmentation. *Biomed. Signal Process. Control* **2021**, *66*, 102481. [CrossRef]
37. Tajbakhsh, N.; Jeyaseelan, L.; Li, Q.; Chiang, J.N.; Wu, Z.; Ding, X. Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation. *Med. Image Anal.* **2020**, *63*, 101693. [CrossRef] [PubMed]
38. Zeiser, F.A.; da Costa, C.A.; Zonta, T.; Marques, N.M.C.; Roehe, A.V.; Moreno, M.; Righi, R.d.R. Segmentation of masses on mammograms using data augmentation and deep learning. *J. Digit. Imaging* **2020**, *33*, 858–868. [CrossRef]
39. Valanarasu, J.M.J.; Oza, P.; Hacihaliloglu, I.; Patel, V.M. Medical transformer: Gated axial-attention for medical image segmentation. In Proceedings of the Medical Image Computing and Computer Assisted Intervention—MICCAI 2021: 24th International Conference, Strasbourg, France, 27 September–1 October 2021; pp. 36–46.
40. Su, Y.; Liu, Q.; Xie, W.; Hu, P. YOLO-LOGO: A transformer-based YOLO segmentation model for breast mass detection and segmentation in digital mammograms. *Comput. Methods Programs Biomed.* **2022**, *221*, 106903. [CrossRef]
41. Cho, P.; Yoon, H.-J. Evaluation of U-net-based image segmentation model to digital mammography. *Med. Imaging Image Process.* **2021**, *11596*, 593–599.
42. Mehta, S.; Mercan, E.; Bartlett, J.; Weaver, D.; Elmore, J.G.; Shapiro, L. Y-Net: Joint segmentation and classification for diagnosis of breast biopsy images. In Proceedings of the Medical Image Computing and Computer Assisted Intervention—MICCAI 2018: 21st International Conference, Granada, Spain, 16–20 September 2018; pp. 893–901.
43. Baccouche, A.; Garcia-Zapirain, B.; Olea, C.C.; Elmaghraby, A.S. Elmaghraby. Connected-UNets: A deep learning architecture for breast mass segmentation. *NPJ Breast Cancer* **2021**, *7*, 151. [CrossRef]
44. Zhu, W.; Xiang, X.; Tran, T.D.; Hager, G.D.; Xie, X. Adversarial deep structured nets for mass segmentation from mammograms. In Proceedings of the 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), Washington, DC, USA, 4–7 April 2018; pp. 847–850.
45. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 3431–3440.
46. Ciresan, D.; Giusti, A.; Gambardella, L.; Schmidhuber, J. Deep neural networks segment neuronal membranes in electron microscopy images. *Adv. Neural Inf. Process. Syst.* **2012**, *25*, 2843–2851.
47. 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 Proceedings of the 2019 IEEE 19th International Conference on Software Quality, Reliability and Security Companion (QRS-C), Sofia, Bulgaria, 22–26 July 2019; pp. 362–367.
48. Zhang, J.; Jin, Y.; Xu, J.; Xu, X.; Zhang, Y. MDU-net: Multi-scale densely connected u-net for biomedical image segmentation. *arXiv* **2018**, arXiv:1812.00352.

49. Li, C.; Tan, Y.; Chen, W.; Luo, X.; Gao, Y.; Jia, X.; Wang, Z. Unet++: A nested u-net architecture for medical image segmentation. In Proceedings of the 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, 20 September 2018; pp. 3–11.
50. 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 Proceedings of the 2020 IEEE International Conference on Image Processing (ICIP), Abu Dhabi, United Arab Emirates, 25–28 October 2020; pp. 345–349.
51. 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 Proceedings of the European Conference on Computer Vision, Tel Aviv, Israel, 23–27 October 2022; pp. 205–218.
52. Oktay, O.; Schlemper, J.; Folgoc, L.L.; Lee, M.; Heinrich, M.; Misawa, K.; Mori, K.; McDonagh, S.; Hammerla, N.Y.; Kainz, B.; et al. Attention u-net: Learning where to look for the pancreas. *arXiv* **2018**, arXiv:1804.03999.
53. 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* **2019**, *7*, 166823–166832. [CrossRef]
54. Pi, J.; Qi, Y.; Lou, M.; Li, X.; Wang, Y.; Xu, C.; Ma, Y. FS-UNet: Mass segmentation in mammograms using an encoder-decoder architecture with feature strengthening. *Comput. Biol. Med.* **2021**, *137*, 104800. [CrossRef]
55. 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 Proceedings of the ICASSP 2020—2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Barcelona, Spain, 4–8 May 2020; pp. 1055–1059.
56. Drozdzal, M.; Vorontsov, E.; Chartrand, G.; Kadoury, S.; Pal, C. The importance of skip connections in biomedical image segmentation. In Proceedings of the International Workshop on Deep Learning in Medical Image Analysis, International Workshop on Large-Scale Annotation of Biomedical Data and Expert Label Synthesis, Granada, Spain, 20 September 2016; pp. 179–187.
57. Zhou, Z.; Siddiquee, M.M.R.; Tajbakhsh, N.; Liang, J. Unet++: Redesigning skip connections to exploit multiscale features in image segmentation. *IEEE Trans. Med. Imaging* **2019**, *39*, 1856–1867. [CrossRef] [PubMed]
58. 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. Eng.* **2019**, *2019*, 8415485. [CrossRef] [PubMed]
59. 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.* **2017**, *40*, 834–848. [CrossRef]
60. Jégou, S.; Drozdzal, M.; Vazquez, D.; Romero, A.; Bengio, Y. The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Honolulu, HI, USA, 21–26 July 2017; pp. 11–19.
61. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4700–4708.
62. Li, S.; Dong, M.; Du, G.; Mu, X. Attention dense-u-net for automatic breast mass segmentation in digital mammogram. *IEEE Access* **2019**, *7*, 59037–59047. [CrossRef]
63. Chen, J.; Chen, L.; Wang, S.; Chen, P. A novel multi-scale adversarial networks for precise segmentation of x-ray breast mass. *IEEE Access* **2020**, *8*, 103772–103781. [CrossRef]
64. Pihur, V.; Datta, S.; Datta, S. Weighted rank aggregation of cluster validation measures: A Monte Carlo cross-entropy approach. *Bioinformatics* **2007**, *23*, 1607–1615. [CrossRef]
65. Lin, T.Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal loss for dense object detection. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2980–2988.
66. Zhu, W.; Huang, Y.; Zeng, L.; Chen, X.; Liu, Y.; Qian, Z.; Du, N.; Fan, W.; Xie, X. AnatomyNet: Deep learning for fast and fully automated whole-volume segmentation of head and neck anatomy. *Med. Phys.* **2019**, *46*, 576–589. [CrossRef]
67. Xie, S.; Tu, Z. Holistically-nested edge detection. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1395–1403.
68. Pan, S.; Zhang, W.; Zhang, W.; Xu, L.; Fan, G.; Gong, J.; Zhang, B.; Gu, H. Diagnostic model of coronary microvascular disease combined with full convolution deep network with balanced cross-entropy cost function. *IEEE Access* **2019**, *7*, 177997–178006. [CrossRef]
69. Milletari, F.; Navab, N.; Ahmadi, S.-A. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In Proceedings of the 2016 Fourth International Conference on 3D Vision (3DV), Stanford, CA, USA, 25–28 October 2016; pp. 565–571.
70. Shruti, J. A survey of loss functions for semantic segmentation. In Proceedings of the 2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB), Fully Virtual, Vina del Mar, Chile, 27–29 October 2020; pp. 1–7.
71. Crum, W.; Camara, O.; Hill, D. Generalized overlap measures for evaluation and validation in medical image analysis. *IEEE Trans. Med. Imaging* **2006**, *25*, 1451–1461. [CrossRef]
72. Fidon, L.; Li, W.; Garcia-Peraza-Herrera, L.C.; Ekanayake, J.; Kitchen, N.; Ourselin, S.; Vercauteren, T. Generalised wasserstein dice score for imbalanced multi-class segmentation using holistic convolutional networks. In Proceedings of the Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: Third International Workshop, Quebec City, QC, Canada, 14 September 2017.

73. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [CrossRef]
74. Aliniya, P.; Razzaghi, P. Parametric and nonparametric context models: A unified approach to scene parsing. *Pattern Recognit.* **2018**, *84*, 165–181. [CrossRef]
75. Alinia, P.; Razzaghi, P. Similarity based context for nonparametric scene parsing. In Proceedings of the 2017 Iranian Conference on Electrical Engineering (ICEE), Tehran, Iran, 2–4 May 2017; pp. 1509–1514.
76. Yeung, M.; Sala, E.; Schönlieb, C.B.; Rundo, L. Focus U-Net: A novel dual attention-gated CNN for polyp segmentation during colonoscopy. *Comput. Biol. Med.* **2021**, *137*, 104815. [CrossRef]