Improving Segmentation of Multiple Instances of Various Sizes in Biomedical Images using Instance-wise and Center-of-Instance Loss Function

Author name(s) withheld EMAIL(S) WITHHELD
Address withheld

Editors: Under Review for MIDL 2023

Abstract

In this paper, we propose a novel two-component loss for biomedical image segmentation tasks called the instance-wise and center-of-instance (ICI) loss, a loss function that addresses the instance imbalance problem commonly encountered when using pixel-wise loss functions such as the Dice loss. The instance-wise component improves the detection of small instances or "blobs" in image datasets with both large and small instances. The center-of-instance component improves the overall detection accuracy. We compared the ICI loss with two existing losses, the Dice loss and the blob loss, in the task of stroke lesion segmentation using the ATLAS R2.0 challenge dataset from MICCAI 2022. Compared to the other losses, the ICI loss provided a better balanced segmentation, and significantly outperformed the Dice loss with an improvement of $1.7 - 3.7\%$ and the blob loss by $0.6 - 5.0\%$ in terms of the Dice similarity coefficient on both validation and test set, suggesting that the ICI loss is a potential solution to the instance imbalance problem.

Keywords: Instance-wise segmentation loss, center-of-instance segmentation loss.

1. Introduction

Object segmentation in biomedical images is a common task, yet presents challenges namely class imbalance and instance imbalance problems due to the diversity of object sizes. Class imbalance problem happens when the number of pixels of a class is much higher than the other classes. Whereas, instance imbalance problem happens when larger instances dominates over smaller instances of the same class. These are two frequent problems that arise when objects appear as multiple, diversely-sized instances in an image, such as stroke lesions in brain magnetic resonance imaging (MRI) (Guerrero et al., 2018). Addressing these challenges is crucial for accurate segmentation and improved diagnostic results.

Deep learning semantic segmentation models often utilize a pixel-wise loss function to evaluate the quality of the segmentations produced by the model across the whole image. Previous studies have demonstrated that pixel-wise loss functions such as cross-entropy (CE) and Dice losses are effective at segmenting large objects and instances (Ronneberger et al., 2015; Milletari et al., 2016). However, pixel-wise loss functions have difficulty in identifying instances of varying sizes, as they tend to miss small features within large instances, and fail to detect individual small instances (Maulana et al., 2021). This is mainly due to their focus on individual pixels, rather than considering the context of objects in an image (Reinke et al., 2021).

Recent advancements in biomedical image analysis have led to the development of new loss functions (Ma et al., 2021), e.g. Tversky loss (Salehi et al., 2017), Focal loss (Lin et al.,

© 2023 Anonymous.

2017), Generalized Dice loss (Sudre et al., 2017), Dice loss with TopK (Brugnara et al., 2020), a combination of Dice loss with CE loss (Isensee et al., 2021), and the Boundary loss (Kervadec et al., 2019), all designed to address the class imbalance problem. However, most of these methods are pixel-wise and fail to tackle the instance imbalance problem.

The instance imbalance problem, where larger instances dominate smaller instances, remains a significant challenge. A normalized instance-wise loss function, where a loss value is computed for each instance individually, is required to address this problem. In this study, **our main contribution is the proposal of a novel instance-wise compound loss function, named the instance-wise and center-of-instance (ICI) loss function**, which can be utilized in conjunction with any pixel-wise loss function for regularization. Through experimental evaluation, we demonstrate the superior performance of the ICI loss function compared to other related loss functions.

2. Related Approaches

Several loss functions have been proposed to solve the instance imbalance problem in biomedical image segmentation, including inverse weighting (Shirokikh et al., 2020), the blob loss (Kofler et al., 2022), and the lesion-wise loss (Zhang et al., 2021).

Inverse weighting (IW) was proposed as an instance-weighted loss function, where each instance is given a weight that is inversely proportional to its size by using a global weight map (Shirokikh et al., 2020). However, IW is implemented by assigning weights to each individual pixel, and is not computed on each instance separately.

In contrast, the **blob loss** is an instance-wise loss calculated for each instance in the label and averaged over all instances (Kofler et al., 2022). But the blob loss is overly sensitive to false segmentations as it includes false segmentations from other instances, even if they do not intersect with the instance-of-interest (see Figure 1 for 2D visualization). This is because connected component analysis (CCA) is precomputed outside of the blob loss and only for the label image, so the blob loss cannot distinguish which instances of the output image intersect with each instance of the output image.

Lastly, the **Lesion-wise loss (LesLoss)** was proposed to assign each blob the same size by transforming all instances into spheres with a fixed size based on instances' centers of mass (Zhang et al., 2021). Similar to the blob loss, the CCA is precomputed outside of the LesLoss and only for the label image. Also, an additional segmentation network is needed to perform segmentation of instances' spheres which limits its applicability.

3. Proposed Approach

Our proposed **Instance-wise and Center-of-Instance (ICI) loss function** is inspired by both the blob loss (Kofler et al., 2022) and LesLoss (Zhang et al., 2021), where we combine instance-wise loss calculation with the normalization of all instances into a square (for 2D images) or a cube (for 3D images) of a fixed size. The ICI loss function consists of two loss calculations: the **instance-wise loss** and the **center-of-instance loss** calculations. In general, the instance-wise loss, distinct from the existing blob loss (Kofler et al., 2022), is used to minimize missed segmentations of instances in the label image (false negatives), especially the small instances. Whereas, the center-of-instance loss is used to improve the

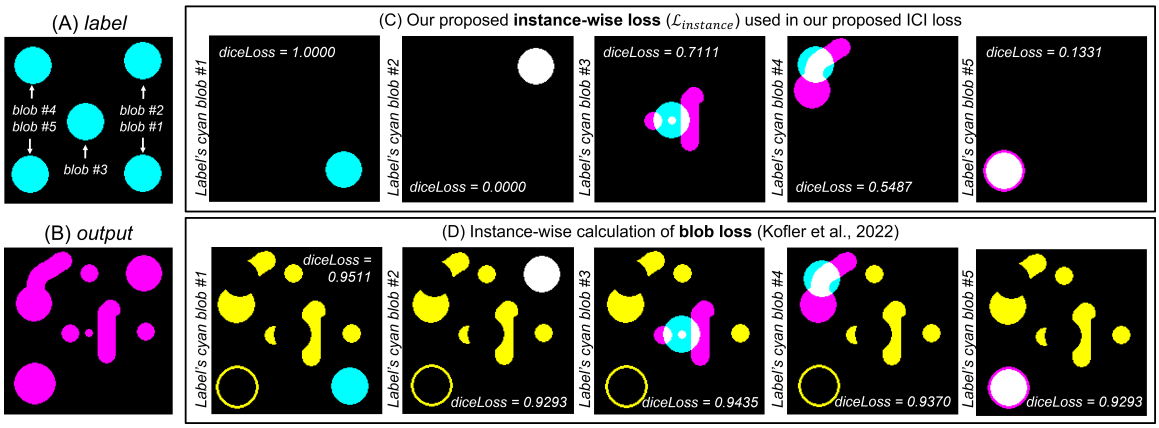

Figure 1: Comparison of instance-wise segmentation losses performed by our proposed instance-wise loss ($\mathcal{L}_{instance}$) (C) and the blob loss (Kofler et al., 2022) (D). Artificial data are used for clearer visualization where (A) shows the *label* image, consisting of 5 instances (cyan), and (B) shows *output* image with a lot more instances (magenta). Each column in (C) and (D) represents each instance from the *label* image. In (C), our proposed instance-wise loss only includes false segmentations from *output* instances that have intersections with the instance-of-interest (magenta). In contrast, blob loss (D) includes false segmentations from other *output* instances (yellow pixels). White pixels are true positive segmentations.

segmentation of small instances in the label image (reduce false negatives) and suppress the detection of small and spurious instances in the output image (reduce false positives).

In this study, we used a compound loss which combines the global Dice loss ($\mathcal{L}_{global}$), instance-wise loss ($\mathcal{L}_{instance}$), and center-of-instance loss ($\mathcal{L}_{center}$) with weights a, b, and c, respectively. The compound loss is shown in Equation (1). The Dice loss is also used to calculate the individual components of the ICI loss; the Dice loss equation is shown in Equation (2). The general flow chart and visualization of the ICI loss is shown in Figure 2.

$$\mathcal{L} = a \times \mathcal{L}_{global} + b \times \mathcal{L}_{instance} + c \times \mathcal{L}_{center} \tag{1}$$

$$diceLoss = 1 - \frac{2|y \cap y_{\text{pred}}|}{|y| + |y_{\text{pred}}|} = 1 - \frac{2\,TP}{2\,TP + FP + FN} \tag{2}$$

3.1. Instance-wise loss

Our proposed **instance-wise loss ($\mathcal{L}_{instance}$)** calculation is similar to the blob loss, where an instance-wise segmentation loss is performed by computing the Dice loss for each instance in the manual label. The difference between our instance-wise loss calculation and the blob loss is that CCA is computed on the fly in our proposed loss function (on GPUs) for both the manual labels and the predicted segmentations. By performing CCA on the fly for both labels and predictions, all instances in the predicted segmentation that overlap by at least 1 pixel/voxel with manually labelled instances can be identified. This approach is not performed in the blob loss, which leads to the inclusion of false segmentations that do not have any overlap with the instance (in the manual label) being calculated (see Figure 1).

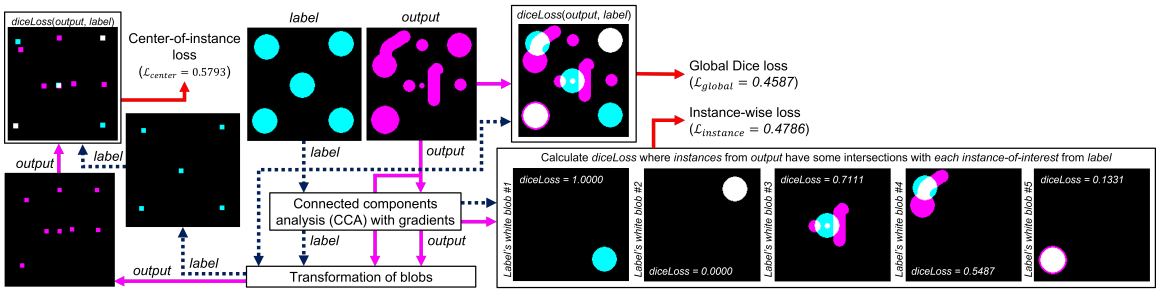

Figure 2: Flow chart and visualization of all losess used in this study. The total loss ($\mathcal{L}$) used in this study is a compound loss, as formulated in Equation (1), consisting of the global Dice loss ($\mathcal{L}_{global}$) and our proposed ICI loss which are instance-wise loss ($\mathcal{L}_{instance}$) and center-of-instance loss ($\mathcal{L}_{center}$).

Instance-wise loss calculation consists of two steps: 1) performing CCA for both the manual label (*label*) and predicted segmentation produced by a deep learning model (*output*), and 2) calculating the Dice loss for each instance in the *label* (instance-of-interest) against any instances in the *output* that intersect with the instance-of-interest (instance-wise Dice segmentation loss). We implemented CCA by modifying the *connected_components* function from the kornia library (Riba et al., 2020) so that all gradients in the predicted segmentation are tracked for backpropagation. In this study, the threshold value of 0.5 was heuristically determined and used to threshold the predicted segmentations before performing CCA. After all instances in *label* (i.e. *cc_label*) and *output* (i.e. *cc_output*) are identified, the instance-wise Dice segmentation loss can be computed by following Algorithm 1.

Algorithm 1: Instance-wise Dice segmentation loss ($\mathcal{L}_{instance}$)

Input: Extracted $cc_label = x_1, \ldots, x_n$ and $cc_output = y_1, \ldots, y_m$ from CCA
Output: bwl `/* mean of the instance-wise Dice segmentation loss */`
$bwl \leftarrow 0$
for $i \leftarrow 1$ **to** n **do**
  $lbl_{x_i} \leftarrow$ label image with only x_i blob `/* other blobs are masked out */`
  $\mathbf{z} = [z_1, \ldots, z_p] \leftarrow$ blobs in cc_output that intersect with x_i
  if $\mathbf{z}$ *is not an empty list* **then**
    $out_{\mathbf{z}} \leftarrow$ output image with blobs in $\mathbf{z}$ `/* other blobs are masked out */`
  else
    $out_{\mathbf{z}} \leftarrow$ output image with no blobs `/* all blobs are masked out */`
  end
  $bwl \leftarrow bwl + diceLoss(out_{\mathbf{z}}, lbl_{x_i})$
end
$bwl \leftarrow bwl/n$

3.2. Center-of-Instance loss

Our proposed **center-of-instance loss ($\mathcal{L}_{center}$)** calculation transforms all blobs into squares (2D) or cubes (3D) with a fixed size before the Dice loss is calculated. Squares/cubes are used instead of circles/spheres (used in the LesLoss) to simplify the transformation's calculation. Simple calculation is important because all gradients need to be tracked for backpropagation on GPUs. All instances in *label* and *output* images can be easily transformed into squares/cubes as the center-of-mass for each instance is calculated during CCA. Using the center-of-mass, each instance can then be transformed into a square/cube by assigning 1 to the pixels/voxels that make up the square/cube area. Visualization of the transformation of 2D blobs into 2D squares (7×7 pixels) is shown in Figure 2. Visualization in 3D space can be found in Appendix B. Pseudo-code of the center-of-instance loss calculation is shown in Algorithm 2 in Appendix C. Based on our preliminary experiments, the best size of the fixed-size cubes for an image of original size $192 \times 192 \times 192$ voxels was found to be $7 \times 7 \times 7$ voxels (see Table 5 in Appendix 5).

4. Experimental Settings

All experimental settings used in this study are described below.

Deep learning model: For all experiments, we used a 3D Residual U-Net (Kerfoot et al., 2018) loaded from the MONAI library. Parameters that were used to create the 3D Residual U-Net are shown in Appendix A. We used a sigmoid function with 0.5 as the threshold value in the segmentation layer for binary segmentation.

Tested loss functions: We compared our proposed ICI loss directly to the blob loss (Kofler et al., 2022) and the global Dice loss. We did not perform any comparisons with LesLoss because an additional network is needed for predicting each instance's sphere. For the blob loss, we used the recommended weights, which are $\alpha = 2$ for the (main) global Dice segmentation loss and $\beta = 1$ for the instance-wise segmentation, which also uses the Dice loss. For our proposed compound ICI loss, we tested different weights for the global Dice loss (a), the instance-wise loss (b), and the center-of-instance loss (c).

Dataset: We used the publicly available ATLAS v2.0 challange dataset from the MICCAI 2022 challenge (Liew et al., 2022) available at https://atlas.grand-challenge.org/. ATLAS v2.0 is a large public dataset of T1w stroke brain MRI and manually segmented lesion masks ($N = 1,271$) that is divided into a public training set ($n = 655$), a test set where the lesion masks are hidden ($n = 300$), and a generalizability set, where both T1w and lesions masks are hidden ($n = 316$). Specific to this study, we only used the public training set for training and validation, and the test set for testing the trained models. We did not use the generalizability set because only one model can be submitted to the system per month. In contrast, one submission can be submitted per day for evaluating the test set.

Training: Out of 655 subjects in the public training set from the ATLAS v2.0 challenge dataset, we manually divided it into a training set ($n = 600$) and validation set ($n = 55$). We performed two different experiments: whole image experiments (with image size of $192 \times 192 \times 192$), and patch-based experiments (with patch size of $96 \times 96 \times 96$) to assess the effectiveness of our proposed ICI loss in segmenting 3D images of different sizes. In the whole image experiment, we trained 3D Residual U-Net models for 200 epochs by using a mini-batch of 4 random subjects in every step. In the patch-based experiment, we

trained 3D Residual U-Net models for 600 epochs (with mini-batches of 2 subjects), where 8 patches were randomly extracted from each subject, with a 1:1 ratio for positive (stroke lesions) and negative (non-stroke lesions) labels. Randomized data augmentations were applied, including left/right flipping, rotation, zooming, and intensity scaling and shifting. Subject-wise intensity normalization was performed by using zero mean unit variance. The 3D Residual U-Net was optimized using the Adam optimizer (Kingma and Ba, 2015). The model that produced the best Dice metric in the validation set was used in testing.

Training environments: We conducted our experiments using various NVIDIA GPUs, including rtxa6000, rtxa5000, v100, a100, and rtx8000, with CUDA version 11.7, Pytorch version 1.13.0, and MONAI version 0.9.0.

Testing/inference: We first performed inference on T1w brain MRI in the test set on our computing nodes, and then we submitted the predicted segmentation results to the ATLAS v2.0 challenge's system. Note that only one submission was permitted per day.

Performance measurements: The ATLAS v2.0 challenge produced 4 performance measurements: Dice similarity coefficient (DSC), volume difference, lesion-wise F1 score, and simple lesion count. We also measured the performance of all models in the validation set by using our own performance measurements, which are DSC, missed instances (MI), false instances (FI), subjects with MI, subjects with FI, and subjects without MI & FI. To decide which model performed best, a numeric rank (written inside square brackets [ ]) is given to each performance measurement such that a mean rank for each model can be calculated.

5. Results

In this section, the ↑ symbol means that higher values are better, while the ↓ symbol means that lower values are better. The best value for each column is shown in bold and the second best is underlined.

5.1. Whole image segmentation

Figure 3 shows that compounding a pixel-wise segmentation loss (i.e., Dice loss) with both terms of our proposed ICI loss (i.e., in $a = 1, b = 1, c = 1$ (yellow lines) and $a = 1/4, b = 1/2, c = 1/4$ (purple lines)) have several advantages in training and validation compared to the other losses. First, yellow and purple lines achieved lower Dice losses in fewer training epochs, as shown in Figure 3A. Second, yellow and purple lines produced lower and more stable numbers of missed and false instances than the other losses during validation, as shown in Figure 3B and C. Lastly, yellow and purple lines achieved higher DSC values more quickly than the other losses during validation as shown in Figure 3D. All these observations suggest that ICI loss successfully regularized Dice loss by keeping the number of missed and false instances low, in addition to lowering the Dice loss itself.

Quantitative results for all losses in the validation set produced by using the best models (i.e., with the highest DSC values) can be seen in Table 1. Note that $\mathcal{L}_{center}$ itself ($a = 1, b = 0, c = 1$) produced the lowest total FI, but failed to produced lower total MI and higher DSC. On the other hand, compounding Dice loss with all ICI loss's terms with optimum weights ($a = 1/4, b = 1/2, c = 1/4$) ranked the best by producing the best DSC with mean rank of 2.33. In contrast, blob loss with recommended weights ($\alpha = 2, \beta = 1$) failed to

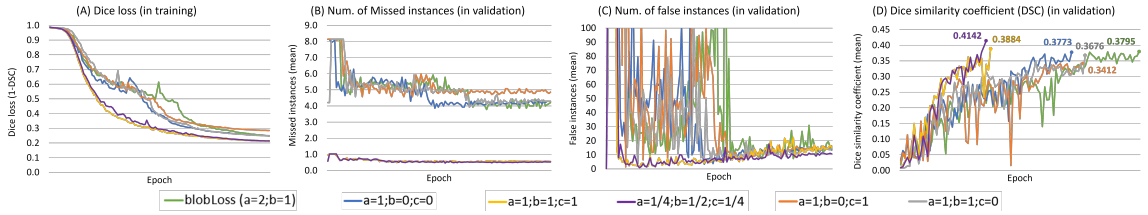

Figure 3: Training curves for (A) Dice loss and validation curves for (B) number of missed blobs, (C) number of false blobs, and (D) DSC values.

Table 1: Performance values on the validation set from the whole image experiments.

Weights (a=global, b=blob, c=center)	Mean Rank (↓)	DSC (↑)	Total MI (mean) (↓)	Subjects w/ MI (↓)	Total FI (mean) (↓)	Subjects w/ FI (↓)	Subjects wo/ MI & FI (↑)	Best at Epoch
blob loss ($\alpha = 2, \beta = 1$)	3.83	0.3795 [3]	20 (0.3636) [2]	16 [2]	382 (6.9455) [6]	53 [5]	4 [5]	107
a = 1, b = 0, c = 0	3.50	0.3773 [4]	23 (0.4182) [5]	19 [4]	149 (2.7091) [3]	46 [2]	12 [3]	77
a = 1, b = 0, c = 1	3.33	0.3412 [6]	29 (0.5273) [6]	21 [5]	**110 (2.0000) [1]**	**37 [1]**	**18 [1]**	82
a = 1, b = 1, c = 0	3.83	0.3676 [5]	21 (0.3818) [3]	17 [3]	205 (3.7273) [4]	51 [4]	9 [4]	83
a = 1, b = 1, c = 1	3.50	0.3884 [2]	**18 (0.3273) [1]**	**15 [1]**	225 (4.0909) [5]	54 [6]	3 [6]	41
a = 1/4, b = 1/2, c = 1/4	**2.33**	**0.4142 [1]**	21 (0.3818) [3]	17 [3]	124 (2.2545) [2]	49 [3]	14 [2]	39

Table 2: Performance values on the test set from the whole image experiments.

Weights (a=global, b=blob, c=center)	Mean Rank (↓)	DSC (↑)	Volume Difference (↓)	Lesion-wise F1 Score (↑)	Simple Lesion Count (↓)
blob loss ($\alpha = 2, \beta = 1$)	5.50	0.3954 (0.3058) [5]	18,358.73 (34,566.44) [6]	0.3036 (0.2374) [5]	6.7900 (29.7589) [6]
$a = 1, b = 0, c = 0$	2.50	0.4147 (0.3097) [4]	16,437.22 (28,680.58) [3]	0.3558 (0.2638) [2]	**4.5367 (6.1965) [1]**
$a = 1, b = 0, c = 1$	4.25	0.3904 (0.3209) [6]	16,148.17 (29,651.68) [2]	0.3028 (0.2579) [6]	4.9967 (6.7012) [5]
$a = 1, b = 1, c = 0$	3.75	0.4160 (0.2991) [3]	17,782.27 (34,580.36) [5]	0.3233 (0.2334) [4]	4.7367 (6.1601) [3]
$a = 1, b = 1, c = 1$	3.25	0.4240 (0.3067) [2]	16,717.15 (27,592.10) [4]	0.3508 (0.2373) [3]	4.9667 (6.4584) [4]
$a = 1/4, b = 1/2, c = 1/4$	**1.25**	**0.4455 (0.3099) [1]**	**15,993.67 (30,064.58) [1]**	**0.4147 (0.2731) [1]**	4.6900 (6.8963) [2]

produce better results except for MI in the validation set, showing a mean rank of 3.83, which is worse than the baseline Dice loss without regularization ($a = 1, b = 0, c = 0$) which showed a mean rank of 3.50.

Table 2 shows that compounding Dice loss with all ICI loss's terms with optimum weights ($a = 1/4, b = 1/2, c = 1/4$) is quite robust to the unseen test set by producing the best values for DSC, Volume Difference, and Lesion-wise F1 Score, and the second best value for Simple Lesion Count with the highest rank (mean rank of 1.25). In contrast, blob loss with recommended weights ($\alpha = 2, \beta = 1$) failed again to achieve a better mean rank than the baseline Dice loss without regularization ($a = 1, b = 0, c = 0$).

5.2. Patch-based segmentation

Our proposed ICI loss also showed superior performance in the patch-based segmentation task when compared to both the baseline Dice loss without regularization ($a = 1, b = 0, c = 0$) and the blob loss with recommended weights ($\alpha = 2, \beta = 1$) in both the validation and test sets, as seen in Tables 3 and 4. However, note that the optimum weights of the ICI loss used in the whole image segmentation experiments (i.e., $a = 1/4, b = 1/2, c = 1/4$) were not the best in terms of mean rank in both validation and test sets, suggesting that the optimal weights for the ICI loss may be dependent on the input image size and task. Nevertheless, Table 3 and Table 4 show that the ICI loss is robust to the unseen test set, as different weights of the ICI loss consistently achieved higher mean ranks compared to the baseline Dice loss without regularization in both validation and test sets.

Table 3: Performance values on the validation set from the patch-based experiments.

Weights (a=global, b=blob, c=center)	Mean Rank (↓)	DSC (↑)	Total MI (mean) (↓)	Subjects w/ MI (↓)	Total FI (mean) (↓)	Subjects w/ FI (↓)	Subjects wo/ MI & FI (↑)	Best at Epoch
blob loss ($\alpha = 2, \beta = 1$)	2.83	0.5237 [3]	**5 (0.0909) [1]**	**5 [1]**	491 (8.9273) [5]	54 [4]	6 [3]	403
a = 1, b = 0, c = 0	3.67	0.5124 [5]	**5 (0.0909) [1]**	**5 [1]**	510 (9.2727) [6]	55 [5]	4 [4]	292
a = 1, b = 0, c = 1	3.00	0.5082 [6]	8 (0.1455) [3]	8 [3]	468 (8.5091) [3]	**50 [1]**	7 [2]	387
a = 1, b = 1, c = 0	**2.00**	**0.5310 [1]**	9 (0.1636) [4]	8 [3]	446 (8.1091) [2]	**50 [1]**	**8 [1]**	440
a = 1, b = 1, c = 1	2.67	0.5201 [4]	8 (0.1455) [3]	8 [3]	**434 (7.8909) [1]**	51 [2]	6 [3]	422
a = 1/4, b = 1/2, c = 1/4	2.33	0.5295 [2]	7 (0.1273) [2]	7 [2]	480 (8.7273) [4]	52 [3]	**8 [1]**	422

Table 4: Performance values on the test set from the patch-based experiments.

Weights (a=global, b=blob, c=center)	Mean Rank (↓)	DSC (↑)	Volume Difference (↓)	Lesion-wise F1 Score (↑)	Simple Lesion Count (↓)
blob loss ($\alpha = 2, \beta = 1$)	5.25	0.5754 (0.2743) [4]	12,210.94 (24,185.93) [5]	0.4033 (0.2413) [6]	5.9933 (7.9349) [6]
$a = 1, b = 0, c = 0$	5.00	0.5598 (0.2724) [6]	13,857.66 (28,911.29) [6]	0.4092 (0.2497) [5]	4.9900 (6.2128) [3]
$a = 1, b = 0, c = 1$	4.00	0.5713 (0.2764) [5]	12,031.63 (24,436.64) [4]	0.4263 (0.2633) [3]	5.3000 (6.6907) [4]
$a = 1, b = 1, c = 0$	2.50	0.5805 (0.2822) [2]	**10,978.48 (22,846.95) [1]**	0.4314 (0.2621) [2]	5.3800 (6.7756) [5]
$a = 1, b = 1, c = 1$	**1.75**	0.5790 (0.2717) [3]	11,564.10 (23,959.23) [2]	**0.4480 (0.2557) [1]**	**4.7800 (6.5443) [1]**
$a = 1/4, b = 1/2, c = 1/4$	2.50	**0.5817 (0.2705) [1]**	11,588.88 (23,726.86) [3]	0.4256 (0.2362) [4]	4.9867 (6.1807) [2]

6. Conclusion

This paper presents a novel Instance-wise and Center-of-Instance (ICI) loss which improved the segmentation of multiple instances with various sizes in biomedical images. In this study, we compared our ICI loss with the Dice loss, a popular pixel-wise segmentation loss, and the blob loss, which was proposed as an instance-wise segmentation loss, in the task of stroke lesion segmentation on the ATLAS R2.0 challenge dataset from MICCAI 2022. Our experiments show that using the ICI loss led to an average increase of 2.7% in segmentation accuracy compared to the Dice loss and 2.4% compared to the blob loss in both whole image segmentation and patch-based segmentation. The code/implementation of ICI loss in Pytorch is available at (withheld).

There are many applications in biomedical image analysis in which the ICI loss may be useful, because many objects that are common targets of segmentation tasks consist of multiple instances of various sizes. The ICI loss has similar limitations to the blob loss: specifically, additional computational resources are required for performing CCA, and performance of the loss function may be sensitive to the weights and hyperparameters used. However, our experiments have shown that a simple set of weights $a = 1, b = 1, c = 1$, without extensive hyperparameter tuning, is sufficient to improve segmentation results in all of the cases based on the DSC metric. The next step is to evaluate whether the ICI loss performs well in multi-class segmentation problems where some classes present as multiple instances with various sizes while others do not. Furthermore, combination with other pixel-wise losses such as CE and Boundary losses might be explored in future studies.

Acknowledgments

Acknowledgments withheld.

References

Gianluca Brugnara, Fabian Isensee, Ulf Neuberger, David Bonekamp, Jens Petersen, Ricarda Diem, Brigitte Wildemann, Sabine Heiland, Wolfgang Wick, Martin Bendszus, et al. Automated volumetric assessment with artificial neural networks might enable a

more accurate assessment of disease burden in patients with multiple sclerosis. *European Radiology*, 30(4):2356–2364, 2020.

Ricardo Guerrero, Chen Qin, Ozan Oktay, Christopher Bowles, Liang Chen, Richard Joules, Robin Wolz, M del C Valdés-Hernández, David Alexander Dickie, Joanna Wardlaw, et al. White matter hyperintensity and stroke lesion segmentation and differentiation using convolutional neural networks. *NeuroImage: Clinical*, 17:918–934, 2018.

Fabian Isensee, Paul F Jaeger, Simon AA Kohl, Jens Petersen, and Klaus H Maier-Hein. nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature methods*, 18(2):203–211, 2021.

Eric Kerfoot, James Clough, Ilkay Oksuz, Jack Lee, Andrew P King, and Julia A Schnabel. Left-ventricle quantification using residual u-net. In *International Workshop on Statistical Atlases and Computational Models of the Heart*, pages 371–380. Springer, 2018.

Hoel Kervadec, Jihene Bouchtiba, Christian Desrosiers, Eric Granger, Jose Dolz, and Ismail Ben Ayed. Boundary loss for highly unbalanced segmentation. In *International conference on medical imaging with deep learning*, pages 285–296. PMLR, 2019.

Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In Yoshua Bengio and Yann LeCun, editors, *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, 2015. URL http://arxiv.org/abs/1412.6980.

Florian Kofler, Suprosanna Shit, Ivan Ezhov, Lucas Fidon, Izabela Horvath, Rami Al-Maskari, Hongwei Li, Harsharan Bhatia, Timo Loehr, Marie Piraud, et al. blob loss: instance imbalance aware loss functions for semantic segmentation. Eprint arXiv:2205.08209, 2022.

Sook-Lei Liew, Bethany P Lo, Miranda R Donnelly, Artemis Zavaliangos-Petropulu, Jessica N Jeong, Giuseppe Barisano, Alexandre Hutton, Julia P Simon, Julia M Juliano, Anisha Suri, et al. A large, curated, open-source stroke neuroimaging dataset to improve lesion segmentation algorithms. *Scientific data*, 9(1):1–12, 2022.

Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, pages 2980–2988, 2017.

Jun Ma, Jianan Chen, Matthew Ng, Rui Huang, Yu Li, Chen Li, Xiaoping Yang, and Anne L Martel. Loss odyssey in medical image segmentation. *Medical Image Analysis*, 71:102035, 2021.

Rizal Maulana, Muhammad Febrian Rachmadi, and Laksmita Rahadianti. Robustness of probabilistic u-net for automated segmentation of white matter hyperintensities in different datasets of brain mri. In *2021 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*, pages 1–7. IEEE, 2021.

Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *2016 fourth international conference on 3D vision (3DV)*, pages 565–571. IEEE, 2016.

Annika Reinke, Matthias Eisenmann, Minu Dietlinde Tizabi, Carole H Sudre, Tim Rädsch, Michela Antonelli, Tal Arbel, Spyridon Bakas, M Jorge Cardoso, Veronika Cheplygina, et al. Common limitations of performance metrics in biomedical image analysis. 2021.

Edgar Riba, Dmytro Mishkin, Daniel Ponsa, Ethan Rublee, and Gary Bradski. Kornia: an open source differentiable computer vision library for pytorch. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 3674–3683, 2020.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

Seyed Sadegh Mohseni Salehi, Deniz Erdogmus, and Ali Gholipour. Tversky loss function for image segmentation using 3d fully convolutional deep networks. In *International workshop on machine learning in medical imaging*, pages 379–387. Springer, 2017.

Boris Shirokikh, Alexey Shevtsov, Anvar Kurmukov, Alexandra Dalechina, Egor Krivov, Valery Kostjuchenko, Andrey Golanov, and Mikhail Belyaev. Universal loss reweighting to balance lesion size inequality in 3d medical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 523–532. Springer, 2020.

Carole H Sudre, Wenqi Li, Tom Vercauteren, Sebastien Ourselin, and M Jorge Cardoso. Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. In *Deep learning in medical image analysis and multimodal learning for clinical decision support*, pages 240–248. Springer, 2017.

Hang Zhang, Jinwei Zhang, Chao Li, Elizabeth M Sweeney, Pascal Spincemaille, Thanh D Nguyen, Susan A Gauthier, Yi Wang, and Melanie Marcille. All-net: Anatomical information lesion-wise loss function integrated into neural network for multiple sclerosis lesion segmentation. *NeuroImage: Clinical*, 32:102854, 2021.

Appendix A. 3D Residual U-Net from MONAI library package

```
from monai.networks.nets import UNet

model = UNet(
    spatial_dims=3,
    in_channels=1,
    out_channels=1,
    channels=(16,32,64,128,256),
    strides=(2, 2, 2, 2),
    num_res_units=2,
    norm=Norm.BATCH,
)
```

Appendix B. Visualization of ICI loss in 3D

Figure 4 visualizes the transformation of instances of stroke lesions into cubes in 3D space for both the label and the prediction. In this visualization, one can appreciate how a small prediction (white blob in the (C) figure) is transformed into a cube of size of $7 \times 7 \times 7$ voxels (white cube in the (D) figure) like all other instances in the image. Note that the size of the transformed cubes/squares can be changed (i.e., the parameter e in this paper, see Algorithm 2 in Appendix C).

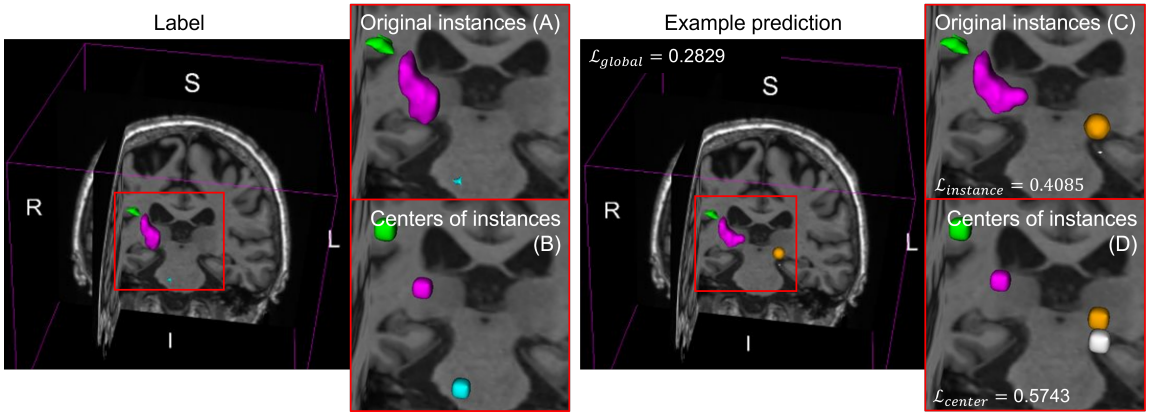

Figure 4: Visualization of ICI loss in 3D with $\mathcal{L}_{global}$ (0.2829), $\mathcal{L}_{instance}$ (0.4085), and $\mathcal{L}_{center}$ (0.5743) values produced by using Dice loss function Equation (2) are shown.

Appendix C. Pseudo-code for center-of-instance loss ($\mathcal{L}_{center}$)

Algorithm 2: Center-of-instance segmentation loss $\mathcal{L}_{center}$

Input: *label* image, *output* image, extracted $cc_{label} = x_1, \ldots, x_n$ from *label* and $cc_{output} = y_1, \ldots, y_m$ from *output* by using CCA, and the new size e for creating square/cube blobs

Output: *cbl* `/* the center-of-instance segmentation loss` $\mathcal{L}_{center}$ `*/`

$label_{cbl} \leftarrow label * 0$ `/* copy label image with all zeros */`

$output_{cbl} \leftarrow output * 0$ `/* copy output image with all zeros */`

for $i \leftarrow 1$ **to** n **do**

  $com \leftarrow$ index of the center of the mass for x_i blob

  $com_e \leftarrow$ indices of the newly transformed com using the input parameter e

  $label_{cbl}(com_e) \leftarrow 1 + \epsilon$ `/* ` ϵ `is a smoothing parameter */`

end

for $i \leftarrow 1$ **to** m **do**

  $com \leftarrow$ index of the center of the mass for y_i blob

  $com_e \leftarrow$ indices of the newly transformed com using the input parameter e

  $output_{cbl}(com_e) \leftarrow 1 + \epsilon$ `/* ` ϵ `is a smoothing parameter */`

end

$cbl \leftarrow diceLoss(output_{cbl}(com_e), label_{cbl}(com_e))$

Appendix D. Results of parameter search for deciding the optimum size of center-of-instance ($\mathcal{L}_{center}$)

Table 5: Validation measurements produced by 3D U-Net models trained for 200 epochs by using full data resolution (i.e., $192 \times 192 \times 192$) and sigmoid function for segmentation (with threshold of 0.5) with different sizes of center-of-instance.

Weights	Size of Center	Mean Rank (↓)	DSC (↑)	Missed Instances (↓)	False Instances (↓)	Losses (↓) Center	Losses (↓) Blob	Best at Epoch
a=1; b=1; c=1	3	3.6	0.3625 [3]	4.2 [3]	9.6 [3]	0.9261 [7]	0.7725 [2]	80
a=1; b=1; c=1	5	4.0	0.3466 [5]	4.9 [5]	**7.4 [1]**	0.9203 [6]	0.7921 [3]	57
a=1; b=1; c=1	7	**2.2**	**0.3884 [1]**	1.9 [2]	8.7 [2]	0.8841 [5]	**0.7532 [1]**	41
a=1; b=1; c=1	11	4.8	0.3437 [6]	4.8 [4]	75.0 [5]	0.8550 [4]	0.8016 [5]	76
a=1; b=1; c=1	15	3.2	0.3689 [2]	**1.8 [1]**	16.6 [6]	0.8487 [3]	0.7936 [4]	50
a=1; b=1; c=1	31	4.2	0.3514 [4]	4.9 [5]	13.6 [4]	0.7253 [2]	0.8258 [6]	84
a=1; b=1; c=1	63	5.2	0.3292 [7]	4.8 [4]	21.5 [7]	**0.6285 [1]**	0.8417 [7]	95

Appendix E. Performance values resulted from test set when softmax function was used for segmentation

Table 6: Performance values on the test set from the whole image segmentation when softmax function was used for segmentation.

Weights (a=global, b=blob, c=center)	Mean Rank (↓)	DSC (↑)	Volume Difference (↓)	Lesion-wise F1 Score (↑)	Simple Lesion Count (↓)
$a = 1, b = 0, c = 0$	**1.25**	0.4056 (0.3142) [2]	**15,341.30 (29,223.21) [1]**	**0.3150 (0.2551) [1]**	**4.8667 (6.5548) [1]**
$a = 1, b = 0, c = 1$	5.00	0.2411 (0.2751) [5]	44,966.23 (39,059.16) [5]	0.1079 (0.0987) [5]	19.4633 (10.7347) [5]
$a = 1, b = 1, c = 0$	3.25	0.3866 (0.2957) [4]	16,776.82 (32,671.45) [3]	0.2155 (0.1890) [3]	9.4567 (8.7481) [3]
$a = 1, b = 1, c = 1$	1.75	**0.4290 (0.2950) [1]**	16,078.24 (32,198.62) [2]	0.2421 (0.1931) [2]	7.7767 (8.2580) [2]
$a = 1/4, b = 1/2, c = 1/4$	3.75	0.3877 (0.2846) [3]	17,868.88 (32,476.77) [4]	0.1787 (0.1519) [4]	14.3200 (13.1372) [4]

Table 7: Performance values on the test set from the patch-based image segmentation when softmax function was used for segmentation.

Weights (a=global, b=blob, c=center)	Mean Rank (↓)	DSC (↑)	Volume Difference (↓)	Lesion-wise F1 Score (↑)	Simple Lesion Count (↓)
$a = 1, b = 0, c = 0$	2.75	0.5770 (0.2795) [5]	12,172.30 (25,327.68) [4]	**0.4453 (0.2545) [1]**	**4.8467 (6.4608) [1]**
$a = 1, b = 0, c = 1$	2.50	0.5811 (0.2702) [3]	11,482.16 (23,263.21) [3]	0.4283 (0.2421) [2]	5.2333 (6.6407) [2]
$a = 1, b = 1, c = 0$	4.25	0.5810 (0.2694) [4]	12,391.04 (25,456.55) [5]	0.4039 (0.2371) [4]	6.0333 (7.3611) [4]
$a = 1, b = 1, c = 1$	**2.00**	**0.5891 (0.2717) [1]**	**11,147.80 (20,211.32) [1]**	0.4170 (0.2511) [3]	5.7600 (6.8380) [3]
$a = 1/4, b = 1/2, c = 1/4$	3.50	0.5858 (0.2718) [2]	11,179.78 (20,233.76) [2]	0.3808 (0.2337) [5]	7.4600 (9.0860) [5]