A Robust Boundary Loss for High Accuracy Medical Image Segmentation

1st Ke Zeng
School of artificial intelligence
Shenzhen Technology University
Shenzhen, China
2410263028@stumail.sztu.edu.cn

2nd Rashid Khan
College of Engineering Physics
Shenzhen Technology University
Shenzhen, China
rashidkhan@sztu.edu.cn

3rd Haomin Liang
School of Artificial Intelligence
Shenzhen Technology University
Shenzhen, China
jamie24618@gmail.com

4th Inayatul Haq
School of Artificial Intelligence
Shenzhen Technology University
Shenzhen, China
inayatulhaq@sztu.edu.cn

5th Zehong Xia
School of Artificial Intelligence
Shenzhen Technology University
Shenzhen, China
zehongxia1@gmail.com

6th Rui Li
School of Artificial Intelligence
Shenzhen Technology University
Shenzhen, China
lirui@sztu.edu.cn

7th Lixin Liang
School of Artificial Intelligence
Shenzhen Technology University
Shenzhen, China
lianglixin@sztu.edu.cn

8th Bingding Huang*
School of Artificial Intelligence
Shenzhen Technology University
Shenzhen, China
huangbingding@sztu.edu.cn

2026 International Joint Conference on Neural Networks (IJCNN) | 979-8-3195-3135-3/26/$31.00 ©2026 IEEE | DOI: 10.1109/IJCNN69854.2026.11728433

***Abstract*—Accurate boundary delineation is critical for medical image segmentation, especially in clinical scenarios involving low contrast, noise, and irregular lesion morphology. Although distance transform-based boundary losses have shown effectiveness in improving geometric accuracy, their optimization often becomes unreliable due to skewed distance distributions and sensitivity to extreme values. This paper proposes Robust Boundary Loss (RBL), an effective loss function designed to enhance optimization reliability and boundary precision. RBL employs a robust normalization strategy for signed distance maps, in which extreme values are suppressed through percentile truncation and informative boundary distances are rescaled into a standardized range. This formulation provides well-behaved and geometrically meaningful gradient signals during optimization, without introducing additional architectural complexity. Extensive experiments on four public datasets spanning three modalities and multiple network architectures demonstrate that RBL consistently reduces boundary errors, yielding clear improvements in 95% Hausdorff Distance (HD95) while maintaining comparable Dice Similarity Coefficient (DSC) and consistent Average Surface Distance (ASD). In addition, RBL exhibits reliable convergence behavior and lower computational overhead compared with Hausdorff-based losses. These results indicate that RBL can serve as a practical and plug-and-play loss function for improving boundary accuracy in challenging medical image segmentation tasks.**

***Index Terms*—Medical image segmentation, boundary-aware loss, robust normalization, distance transform, deep learning**

I. Introduction

In medical image segmentation, precise boundary delineation is vital for clinical diagnosis and treatment planning. For example, in thyroid or breast ultrasound, the regularity and infiltration of lesion margins are key indicators of malignancy (e.g., lobulated or spiculated margins [1]). Similarly, in gastrointestinal endoscopy, accurate boundaries are a prerequisite for determining resection margins in procedures like EMR or ESD [2]. Furthermore, melanoma diagnosis in dermoscopic images relies heavily on morphological irregularities, such as jagged contours, which are core components of the ABCD scoring system [3].

While deep networks like U-Net [4] and its variants [5]–[8] have achieved great success, standard region-based losses (e.g., Dice loss [9] and cross-entropy [10]) often fail to optimize boundary precision. These losses focus on global overlap, making them insensitive to boundary errors that involve only a small fraction of pixels [11]. Consequently, predicted segmentations often exhibit blurred or geometrically inaccurate boundaries. To address this, recent studies have introduced boundary-aware loss functions using geometric priors like distance transforms [12]–[15]. However, two major challenges remain:

1) In low-contrast or noisy images, distance maps often exhibit highly skewed distributions, leading to gradient instability during training.
2) Common normalization strategies (e.g., **global range normalization**) are highly sensitive to outliers, which

This study was supported by the Shenzhen Science and Technology Program (Grant Nos. KJZD20240903095605007 and JCYJ20250604145033042), the Shenzhen Medical Research Fund (No. D250402003), and the Natural Science Foundation of Top Talent of SZTU (Grant No. GDRC202213).

*Corresponding author: huangbingding@sztu.edu.cn.

Authorized licensed use limited to: Shenzhen Technology University. Downloaded on October 26,2026 at 06:26:38 UTC from IEEE Xplore. Restrictions apply.

compresses meaningful distance variations and hinders the model's ability to capture subtle structures.

In this paper, we propose the **Robust Boundary Loss (RBL)** to resolve these stability issues. We observe that signed distance maps in medical image segmentation often exhibit highly imbalanced value distributions, where a small number of extreme distances can dominate the scaling of Boundary Loss and weaken gradient signals near object boundaries during optimization. The core idea of RBL is an outlier-resistant preprocessing mechanism for distance maps. Specifically, RBL suppresses the interference of extreme values through quantile-based truncation and enhances distance discriminability via interval mapping. This design provides stable and geometrically sensitive gradient signals, thereby effectively improving boundary accuracy.

The main contributions of this paper are:

1) A critical analysis of how distance map distributions cause gradient instability and feature compression in existing boundary-aware losses.
2) The RBL function, which incorporates a quantile-based robust normalization strategy. It improves training stability and boundary discriminability while remaining simple to implement.
3) Extensive Validation and Versatility: We conduct comprehensive evaluations across four diverse medical imaging modalities and multiple state-of-the-art architectures. The results demonstrate that RBL consistently improves boundary precision—specifically reducing HD95 by 1-2 points—while ensuring superior training stability and a plug-and-play integration that requires no architectural modifications.

RBL offers a plug-and-play solution for high-precision medical image segmentation, with strong potential for clinical application.

II. Related Work

Medical image segmentation is often challenged by tiny foreground regions, ambiguous boundaries, and complex anatomical structures. While traditional region-based losses, such as Dice loss, effectively handle class imbalance, they lack explicit modeling of geometric structures. Consequently, they often struggle to produce the high fidelity contours required in clinical practice.

Ma et al. [16] provided a comprehensive analysis of various loss functions, revealing that region-based and boundary-based losses have complementary strengths. Their study suggests that no single loss dominates all scenarios, highlighting the need for designs that capture geometric information while ensuring stable optimization.

To improve boundary quality, researchers have explored various geometric and shape priors. For instance, the Generalized Surface Loss (GSL) uses scale normalization to reduce the Hausdorff distance [17], while TopK Dice Loss [18] focuses on the most difficult pixels to enhance sparse region learning. Recently, Inter-CBL [19] introduced an adaptive weighting mechanism to balance category level and sample level contributions, showing strong generalization in kidney and polyp segmentation.

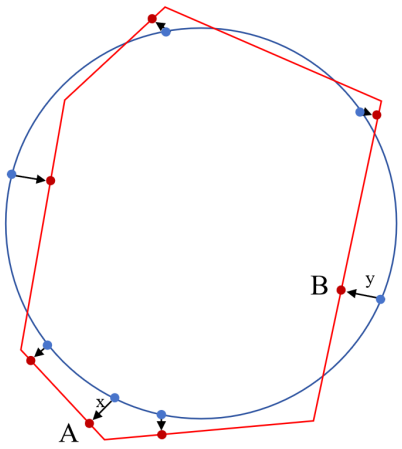


Fig. 1: The blue and red curves denote the ground-truth and predicted boundaries, respectively. For point A, the arrow indicates the shortest distance x to the ground-truth boundary, yielding a positive signed distance $+x$. Conversely, point B has a signed distance of $-y$.

Another research line incorporates boundary metrics or distance representations directly into the optimization. Karimi and Salcudean [20] proposed a loss based on the Hausdorff distance, while others have used Signed Distance Maps (SDMs) [12] or Distance Transform Maps (DTMs) [13] as auxiliary supervision. More recently, FRSLoss [14] applied fuzzy rough set theory to model boundary uncertainty, and FourierLoss [15] utilized Fourier descriptors to constrain contour smoothness in the frequency domain.

Despite their performance, many of these methods are computationally expensive or require complex architectural changes, such as auxiliary branches for additional tasks or online distance map generation. In contrast, the Boundary Loss proposed by Kervadec et al. [21] offers a more elegant and minimalist approach. Boundary Loss guides optimization via a pixel-wise product between the prediction and the ground truth distance function, followed by spatial averaging. Since distance maps can be precomputed offline, Boundary Loss adds negligible training overhead and requires no auxiliary network heads. Its simplicity and effectiveness in imbalanced tasks, like brain tumor segmentation, make it a highly practical baseline for boundary optimization.

III. Method

Boundary Loss [21] has demonstrated strong performance in medical image segmentation by explicitly penalizing boundary errors. Its core idea is to use signed distance maps to quantify boundary proximity. The loss is formulated as the element-wise product between the predicted probability map pc and the signed distance map dc. Here, pc denotes the predicted foreground probability map after sigmoid (binary segmentation setting), and dc's sign convention follows Kervadec et al. [21],

Authorized licensed use limited to: Shenzhen Technology University. Downloaded on October 26,2026 at 06:26:38 UTC from IEEE Xplore. Restrictions apply.

Algorithm 1 Robust Boundary Loss

// Input: predicted probabilities pc and binary mask M

// Compute signed distance map dc:
// $dc < 0$ inside object ($M = 1$), $dc > 0$ outside ($M = 0$)
$dc = \text{COMPUTESIGNEDDISTANCEMAP}(M)$

// Stage 1: Outlier truncation via percentile clipping
$q_1 = \text{PERCENTILE}(dc, 1)$
$q_{99} = \text{PERCENTILE}(dc, 99)$
$dc^{\text{clip}} = \text{CLAMP}(dc,\, q_1,\, q_{99})$

// Stage 2: Robust normalization to $[-1, 1]$
$d_{\min} = \min(dc^{\text{clip}})$
$d_{\max} = \max(dc^{\text{clip}})$
$\hat{dc} = 2 \cdot \dfrac{dc^{\text{clip}} - d_{\min}}{d_{\max} - d_{\min}} - 1$

// Pixel-wise boundary loss
$\mathcal{L} = pc \odot \hat{dc}$

// Final loss averaged over all pixels
$\mathcal{L}_{RBL} = \text{MEAN}(\mathcal{L})$

return $\mathcal{L}_{RBL}$

where distances are negative inside the object and positive outside, as illustrated in Fig. 1.

This operation is performed independently for each pixel, followed by averaging the results across all pixels:

$$\mathcal{L}_{\text{BL}} = \text{MEAN}\left(pc \odot dc\right) \tag{1}$$

where $\odot$ denotes element-wise multiplication. The gradient of this loss with respect to the network parameters is entirely determined by dc. For regularly shaped objects, dc varies smoothly in space, resulting in stable training dynamics. In contrast, when segmenting irregular pathological structures such as tumors or inflammatory regions, dc may exhibit pronounced spatial fluctuations. These fluctuations can induce excessively large gradient magnitudes and, in extreme cases, lead to gradient explosion, thereby causing slow convergence or training instability.

A straightforward mitigation strategy is to apply **global range normalization** to dc, defined as:

$$\bar{dc} = \frac{dc}{dc_{\max} - dc_{\min}} \tag{2}$$

which scales the distance map into a symmetrical $[-1, 1]$ interval. While this approach can stabilize training to some extent, it suffers from a critical limitation when applied to irregular targets. Specifically, due to complex lesion boundaries, the signed distance map typically exhibits a highly skewed distribution, where most pixels concentrate near zero, and only a small number of outliers lie at the extremes. Under such conditions, the global extrema, namely $\max(dc)$ and $\min(dc)$, are dominated by these outliers. As a consequence, distance values corresponding to pixels in the vicinity of object boundaries, where supervision is most informative, are squeezed into a narrow region around the center of the normalized range, resulting in diminished gradient contrast during optimization.

To tackle this issue, we propose a **two-stage robust normalization strategy.** In practice, signed distance maps usually contain extreme values far from object boundaries, which can dominate the loss magnitude and reduce the influence of informative boundary regions. From an optimization perspective, pixels near boundaries contribute most to shape refinement, while distant pixels tend to introduce large yet uninformative gradients. Therefore, our normalization strategy truncates extreme distances and redistributes values adjacent to the boundary over a standardized interval. This suppresses outliers, mitigates gradient noise, and thus supports stable optimization and enhanced boundary discrimination.

Stage 1: Outlier Truncation. We apply percentile-based clipping to the original distance map dc: values below the 1st percentile (q_1) and above the 99th percentile (q_{99}) are truncated to q_1 and q_{99}, respectively, yielding the clipped distance map dc^{clip}. This truncation removes only extreme outliers while largely preserving the relative structure of signed distance maps, thereby reducing the influence of anomalous values on subsequent normalization. An overly aggressive truncation may remove a substantial portion of informative distance values and distort boundary-related geometry, whereas insufficient truncation may fail to adequately suppress extreme outliers and thus provide limited improvement in optimization stability. Therefore, RBL adopts this balanced truncation strategy to simultaneously improve optimization stability and preserve boundary-sensitive distance information.

Stage 2: Robust Normalization over the Clipped Range. We perform normalization only within the effective range of the clipped map and linearly map it to the standard interval $[-1, 1]$, thereby maintaining semantic consistency with the original Boundary Loss. Specifically, let $d_{\min} = \min(dc^{\text{clip}})$ and $d_{\max} = \max(dc^{\text{clip}})$. The normalized distance map is defined as:

$$\hat{dc} = 2 \cdot \frac{dc^{\text{clip}} - d_{\min}}{d_{\max} - d_{\min}} - 1 \tag{3}$$

Compared to the global range normalization ($\bar{dc}$) mentioned above, this approach avoids the dominance of extreme outliers on the scaling factor. Consequently, boundary pixels that were previously compressed near zero are now more uniformly distributed across the full $[-1, 1]$ interval. As a result, boundary-related distance values are expanded to approximately twice their original span under global range normalization, which magnifies gradient signals and enables the model to better capture complex boundary geometries. Finally, we plug the refined distance map $\hat{dc}$ into the Boundary Loss formulation, yielding our proposed loss function:

$$\mathcal{L}_{RBL} = \text{MEAN}\left(pc \odot \hat{dc}\right) \tag{4}$$

Authorized licensed use limited to: Shenzhen Technology University. Downloaded on October 26,2026 at 06:26:38 UTC from IEEE Xplore. Restrictions apply.

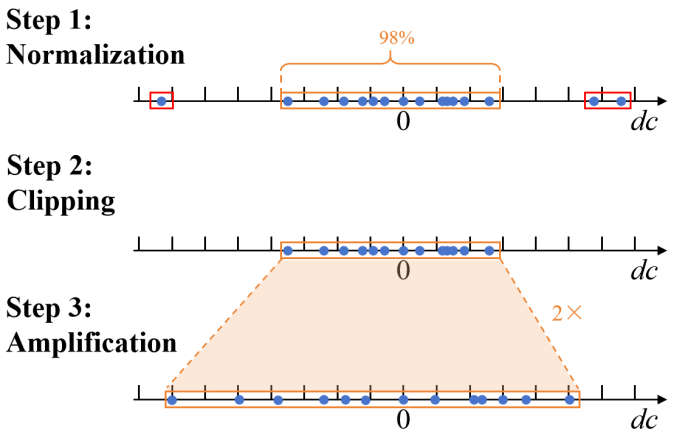


Fig. 2: RBL normalization process.

The normalization process of Robust Boundary Loss's distance map dc is illustrated in Fig. 2 and the workflow is summarized in Algorithm 1.

This design preserves the semantic meaning of the signed distance map while improving both gradient stability and discriminative power through robust normalization, making it particularly suitable for segmenting complex and irregular pathological structures.

IV. Experiments

A. Experimental Environment

All experiments were performed on a Linux based workstation running Ubuntu 20.04.3. The implementation was carried out using Python 3.10.15 and PyTorch 2.0.1 as the deep learning framework. Training and inference were accelerated by an NVIDIA GeForce RTX 3090, supported by CUDA 11.8. To ensure reproducibility, all models were trained and evaluated under identical hardware and software configurations.

B. Datasets

We conducted extensive experiments on four publicly available medical image segmentation datasets, covering diverse imaging modalities (ultrasound, endoscopy, dermoscopy) and anatomical regions (breast, thyroid, colon, skin). All datasets provide pixel level binary masks for lesion segmentation. For fair and consistent evaluation across models, we apply a unified random split of 90% for training and 10% for testing on each dataset, without overlap between splits.

1) **BUSI** [22]: The breast ultrasound image dataset originally contains three classes: benign, malignant, and normal. Following common practice in tumor segmentation, we exclude the "normal" cases and retain only 583 images (benign + malignant) with corresponding binary masks.
2) **TN3K** [23]: A large scale thyroid nodule ultrasound dataset comprising 2,879 images, each annotated with a binary segmentation mask. We use the full set as provided.
3) **CVC-ClinicDB** [24]: A colonoscopy dataset containing 550 high resolution frames of colorectal polyps, each paired with an expert delineated binary ground truth mask.
4) **ISIC 2017** [25]: A dermoscopic skin lesion dataset consisting of 2,750 images from the training and validation partitions of the ISIC 2017 challenge, all labeled with binary masks for lesion regions.

For each dataset, preprocessing (e.g., intensity normalization, resampling) is performed according to the default pipeline recommended by the respective model framework, ensuring compatibility while preserving the integrity of the original data distribution.

C. Experimental Setup

To comprehensively evaluate the effectiveness of the proposed RBL, we conduct systematic experiments across four widely adopted medical image segmentation models: nnU-Net [5], U-Mamba [6], SwinUNet [8], and TransUNet [7]. Each model is independently trained and evaluated on all four datasets: BUSI, TN3K, CVC-ClinicDB, and ISIC 2017.

For U-Mamba, we adopt the "bot" variant (which integrates Mamba blocks only in the bottleneck layer) rather than the more computationally intensive "enc" variant (Mamba blocks in every encoder/decoder stage). Since the loss function operates at the output level and does not interact with internal architecture design, both variants would exhibit identical sensitivity to the loss formulation. Thus, using the bot version provides a faithful yet efficient evaluation of our loss functions.

For every combination of model and dataset, we compare the following five loss configurations:

1) **Baseline**: The default loss function used in the original paper or official implementation of the model.
2) **Baseline + HD Loss**: The baseline augmented with the Hausdorff Distance (HD) loss.
3) **Baseline + Dual HD Loss**: The baseline combined with a symmetric dual formulation of the HD Loss.
4) **Baseline + Boundary Loss**: The baseline enhanced with the classical boundary-aware loss.
5) **Baseline + RBL (Ours)**: The baseline integrated with our proposed RBL.

All experiments are conducted using the default hyperparameters and preprocessing pipelines of each model, with only the loss function component modified to ensure a fair comparison. To mitigate the impact of randomness, each experiment is repeated twice under the same data split, and the average performance is reported.

D. Training Hyperparameters

All models are trained for 400 epochs using the Stochastic Gradient Descent (SGD) optimizer with momentum 0.99, consistent with the default configuration in each model's official implementation. The initial learning rate is set to 0.01 and decayed according to a polynomial schedule:

$$\mathrm{lr}_t = 0.01 \times \left(1 - \frac{t}{400}\right)^{0.9} \tag{5}$$

Authorized licensed use limited to: Shenzhen Technology University. Downloaded on October 26,2026 at 06:26:38 UTC from IEEE Xplore. Restrictions apply.

TABLE I: Performance comparison of different models and loss functions on multiple datasets. Metrics include DSC (%)↑, HD95 (mm)↓, and ASD (mm)↓. Higher DSC and lower HD95/ASD are better. Best value per model per dataset is **bolded**. The "-" indicates that the results are not available because the corresponding experiment failed to converge during training.

Model	Loss	TN3K			BUSI			CVC-ClinicDB			ISIC 2017		
		DSC↑	HD95↓	ASD↓	DSC↑	HD95↓	ASD↓	DSC↑	HD95↓	ASD↓	DSC↑	HD95↓	ASD↓
nnUNet	Baseline	85.86	56.89	15.42	88.65	30.97	10.65	**81.65**	39.08	**10.27**	**83.26**	521.49	178.34
	+ HD Loss	46.40	164.58	82.46	48.03	124.86	50.81	37.91	69.32	25.59	88.95	349.31	113.59
	+ DHD Loss	46.22	162.35	77.69	52.02	110.31	46.01	26.74	68.17	27.15	–	–	–
	+ Boundary Loss	77.96	63.85	21.28	64.25	76.80	27.85	71.20	47.03	13.01	53.73	1700.90	566.78
	+ RBL (ours)	**86.11**	**46.49**	**13.64**	**88.74**	**28.99**	**10.04**	81.40	**38.80**	10.33	83.10	**520.87**	**176.48**
U-Mamba	Baseline (bot)	86.50	42.89	12.27	88.18	29.70	10.52	80.99	40.36	**9.31**	83.40	497.31	**166.42**
	+ HD Loss	30.72	197.65	116.89	–	–	–	29.85	68.88	34.30	–	–	–
	+ DHD Loss	–	–	–	7.60	201.70	116.76	37.82	70.60	30.73	–	–	–
	+ Boundary Loss	78.42	55.14	19.35	75.98	54.14	17.21	69.20	52.17	14.31	78.43	1086.92	359.77
	+ RBL (ours)	**86.84**	**41.30**	**12.18**	**88.49**	**29.14**	**9.65**	**81.25**	**40.03**	10.14	**83.92**	**493.05**	168.61
TransUNet	Baseline	84.11	47.69	13.06	87.35	33.21	11.32	75.18	73.09	19.84	90.29	317.17	89.11
	+ HD Loss	84.32	43.78	**12.32**	88.70	32.10	10.78	77.79	73.33	19.43	**90.72**	**290.97**	84.49
	+ DHD Loss	83.90	46.40	13.42	87.79	31.43	**10.56**	82.09	60.33	**12.52**	90.50	295.83	93.90
	+ Boundary Loss	71.46	70.86	26.53	82.77	43.40	14.30	71.87	63.39	13.04	78.59	643.88	206.31
	+ RBL (ours)	**84.66**	**43.33**	**12.32**	**88.81**	**31.30**	11.42	**82.44**	**59.57**	**13.40**	90.30	294.16	**83.77**
SwinUNet	Baseline	80.15	46.64	13.97	**85.04**	39.56	12.97	69.38	52.24	13.72	87.50	457.17	133.86
	+ HD Loss	80.59	50.11	16.33	84.64	41.14	13.53	56.51	78.97	24.69	89.64	348.36	110.53
	+ DHD Loss	80.90	50.13	15.24	83.96	39.46	13.95	**72.86**	51.58	**12.87**	**89.92**	338.36	107.53
	+ Boundary Loss	72.59	59.17	19.67	69.20	80.99	28.63	59.88	61.99	24.48	–	–	–
	+ RBL (ours)	**80.96**	**45.84**	**13.74**	84.87	**38.16**	**12.75**	65.61	**50.70**	13.39	89.00	**337.22**	**103.24**

where t denotes the current epoch.

The batch sizes are automatically determined by the nnU-Net framework based on image resolution and GPU memory, and then uniformly applied to all models to eliminate variance induced by batch size. Specifically, we used batch sizes of 10 (BUSI), 29 (TN3K), 25 (CVC-ClinicDB), and 4 (ISIC 2017).

For non-baseline configurations, the total loss is formulated as a dynamically weighted combination of the model's default loss $\mathcal{L}_{\text{base}}$ (i.e., the baseline loss used in its original implementation) and a boundary-aware term $\mathcal{L}_{\text{boundary}}$, which corresponds to one of the following: HD Loss, Dual HD Loss, Boundary Loss, or the proposed RBL:

$$\mathcal{L}_{\text{total}} = (1 - \alpha_t) \cdot \mathcal{L}_{\text{base}} + \alpha_t \cdot \mathcal{L}_{\text{boundary}} \tag{6}$$

where $\alpha_t \in [0, 1]$ is an epoch-dependent coefficient that increases linearly with the training epoch t.

$$\alpha_t = \alpha_0 + t \cdot \Delta\alpha, \quad \alpha_0 = 0.01 \tag{7}$$

The increment $\Delta\alpha = 0.001225$ is selected such that $\alpha_t \approx 0.5$ at epoch 400, ensuring that the boundary-aware term gradually becomes influential only after the model has reached a reasonably mature segmentation state. Specifically, during the early training stage, the optimization is dominated by the baseline loss, which facilitates smooth convergence and reliable learning of global shape and region-level structures. As training progresses and the predicted segmentation becomes more coherent, the contribution of the boundary-aware term is progressively increased, enabling finer boundary refinement without compromising optimization reliability. For the Baseline configuration, the total loss degenerates to $\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{base}}$, and no boundary-related supervision is applied.

TABLE II: Training time per epoch (in seconds) on the CVC-ClinicDB dataset under different loss configurations.

Model	Baseline	+ HDLoss	+ DHDLoss	+ BoundaryLoss	+ RBL (ours)
nnUNet	22.4	36.5	89.6	27.9	27.1
U-Mamba	53.2	73.8	191.5	106	106
TransUNet	6.96	7.77	15.33	7.87	7.83
SwinUNet	0.96	9.10	16.70	2.80	2.70

E. Evaluation Metrics

We employ three standard segmentation metrics for quantitative evaluation: Dice Similarity Coefficient (DSC), which measures the overlap between prediction and ground truth; 95% Hausdorff Distance (HD95), which reflects the worst-case boundary deviation; and Average Surface Distance (ASD), which computes the mean distance between the predicted and true boundaries.

Qualitative assessment is performed through visual comparison of predicted masks against ground truth annotations.

V. Results and Analysis

We conduct a comprehensive evaluation of the proposed method across four representative segmentation architectures: nnU-Net, U-Mamba, TransUNet, and SwinUNet. Our experiments cover four diverse medical imaging tasks: TN3K for thyroid ultrasound, BUSI for breast ultrasound, CVC-ClinicDB for colonoscopy polyps, and ISIC 2017 for dermoscopic skin lesions. Following the experimental setup, RBL is

Authorized licensed use limited to: Shenzhen Technology University. Downloaded on October 26,2026 at 06:26:38 UTC from IEEE Xplore. Restrictions apply.

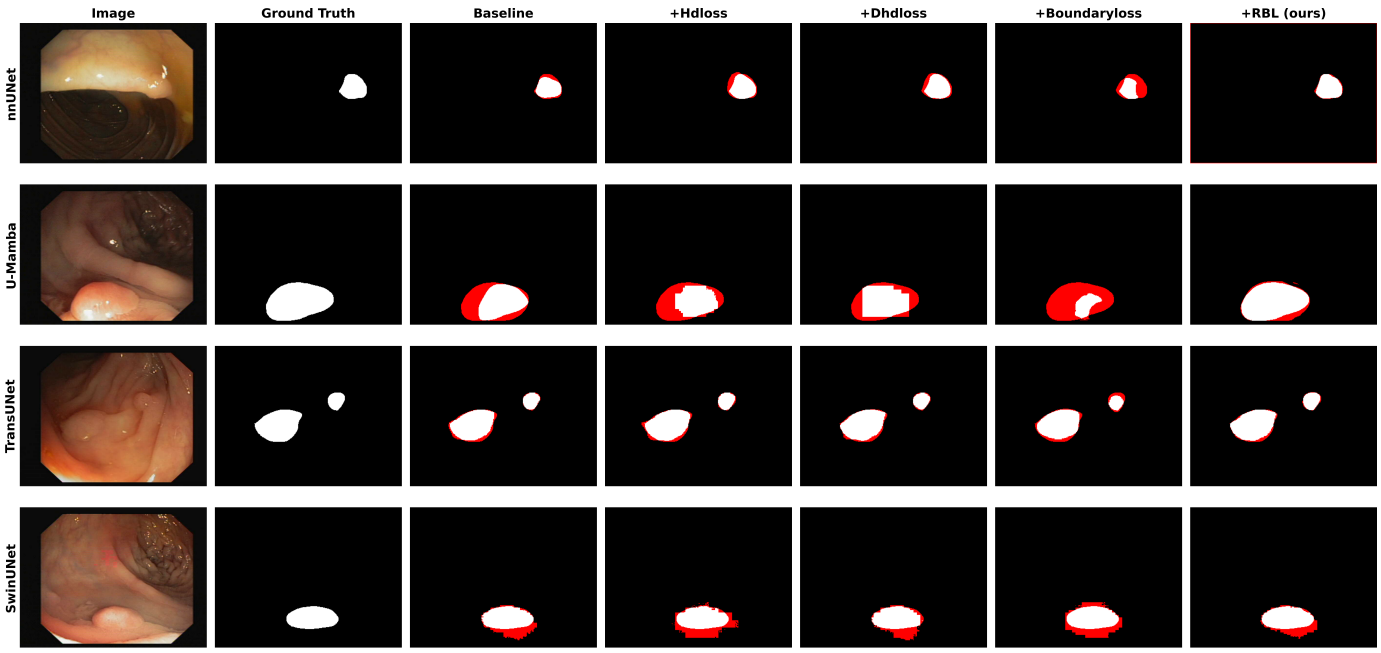


Fig. 3: Qualitative results of four models under five loss functions on the CVC-ClinicDB dataset. Red areas indicate the discrepancies between the ground truth and predictions.

TABLE III: Ablation study on different normalization strategies across multiple medical segmentation datasets. The metrics include DSC (%), HD95 (mm), and ASD (mm). Higher DSC and lower HD95/ASD are better. Each row corresponds to a model variant with one normalization method enabled (✓) and others disabled (×). Best value per model per dataset is **bolded**. The "-" indicates that the results are not avaliable because corresponding experiment failed to converge during training.

Model	Normalization Strategy			TN3K			BUSI			CVC-ClinicDB			ISIC 2017		
	No Norm	Global Range	RBL	DSC↑	HD95↓	ASD↓	DSC↑	HD95↓	ASD↓	DSC↑	HD95↓	ASD↓	DSC↑	HD95↓	ASD↓
nnUNet	✓	×	×	77.96	63.85	21.28	64.25	76.80	27.85	71.20	47.03	13.01	53.73	1700.90	566.78
	×	✓	×	85.98	48.33	14.58	88.46	31.29	10.66	81.00	40.84	10.33	81.42	**536.11**	197.09
	×	×	✓	**86.11**	**46.49**	**13.64**	**88.74**	**28.99**	**10.04**	**81.40**	**38.80**	**10.13**	**82.10**	**520.87**	**176.48**
U-Mamba (bot)	✓	×	×	78.42	55.14	19.35	75.98	54.14	17.21	69.20	52.17	14.31	78.43	1086.92	359.77
	×	✓	×	86.60	43.52	12.33	88.33	30.69	10.25	81.17	40.56	**9.06**	83.84	495.27	170.61
	×	×	✓	**86.84**	**41.30**	**12.18**	**88.49**	**29.14**	**9.65**	**81.25**	**40.03**	10.14	**83.92**	**493.05**	**168.61**
TransUNet	✓	×	×	71.46	70.86	26.53	82.77	43.40	14.30	71.87	63.39	13.04	78.59	643.88	206.31
	×	✓	×	84.22	43.46	**12.23**	87.74	38.77	11.67	78.68	62.65	15.94	**90.38**	317.47	92.03
	×	×	✓	**84.66**	**43.33**	12.32	**88.81**	**31.30**	**11.42**	**82.44**	**59.57**	**12.40**	90.30	**294.16**	**83.77**
SwinUNet	✓	×	×	72.59	59.17	19.67	69.20	80.99	28.63	59.88	61.99	24.48	–	–	–
	×	✓	×	77.22	55.46	17.23	83.07	45.33	15.27	59.16	59.81	17.48	88.12	390.69	128.72
	×	×	✓	**78.96**	**54.84**	**16.74**	**83.27**	**44.96**	**14.75**	**65.61**	**58.70**	**16.39**	**89.00**	**340.22**	**118.24**

evaluated against baseline training and three commonly used boundary-aware loss functions, including the HD Loss, Dual HD Loss, and the original Boundary Loss.

Results reported in Table I show that RBL yields superior geometric accuracy, particularly regarding HD95 and ASD, while preserving competitive DSC across all model-dataset combinations. The most prominent improvements occur on the two ultrasound-based datasets. In these scenarios, our approach proves highly effective at handling low-contrast boundaries and the speckle noise typical of such imaging. These observations suggest that RBL may be particularly beneficial for ultrasound segmentation, where precise delineation is often challenged by imaging artifacts. By providing a more balanced gradient signal near object boundaries through a robust distance weighting mechanism, the algorithm helps models produce anatomically plausible contours without necessitating architectural modifications.

On the CVC-ClinicDB dataset, clear gains in contour smoothness and spatial alignment are also observed. However, the magnitude of improvement is more modest compared to ultrasound scans, likely due to the higher contrast and clearer lesion boundaries inherent in endoscopic images. Similarly, on

Authorized licensed use limited to: Shenzhen Technology University. Downloaded on October 26,2026 at 06:26:38 UTC from IEEE Xplore. Restrictions apply.

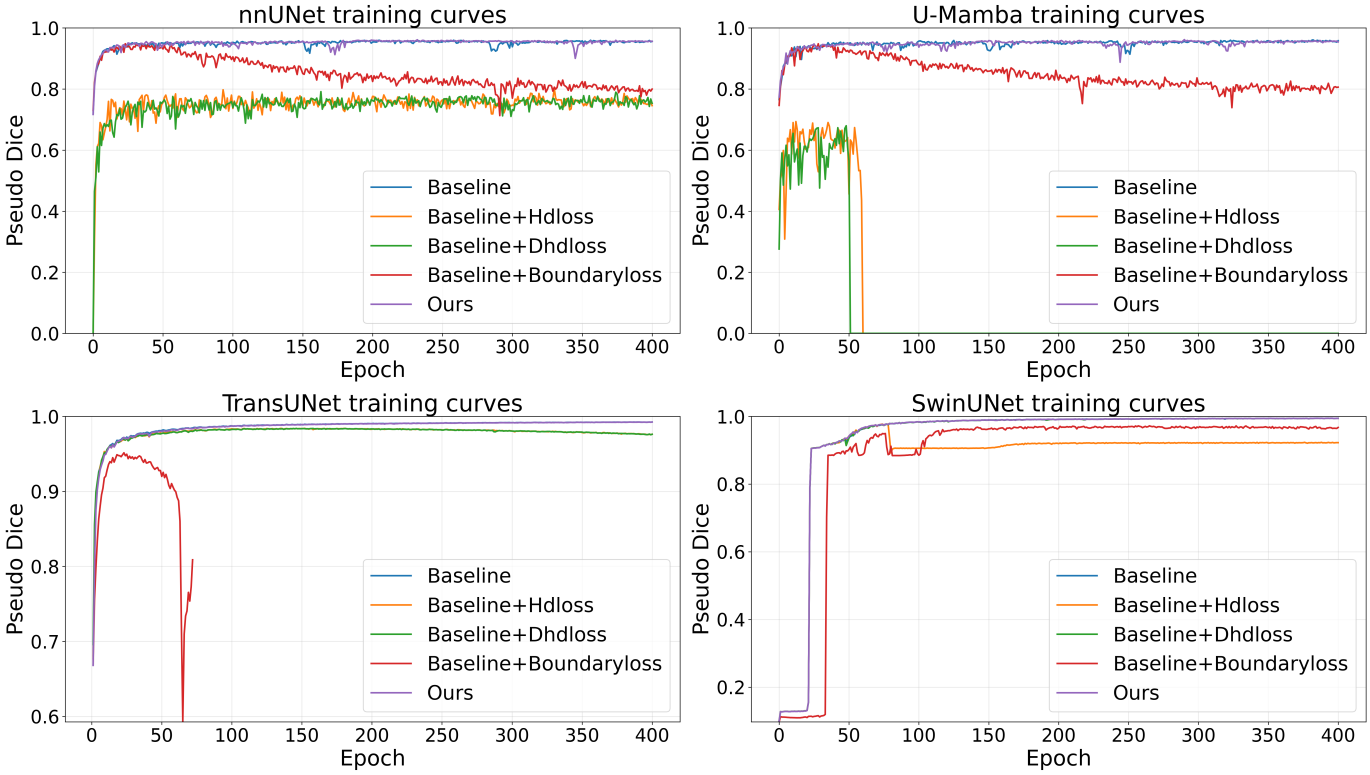


Fig. 4: Four training curves of nnU-Net, U-Mamba, TransUNet, and SwinUNet on CVC-ClinicDB dataset.

ISIC 2017, the Strategy enhances boundary fidelity despite the complex morphology and scale variation of skin lesions, albeit with relatively smaller incremental gains. Taken together, these results indicate that RBL offers a promising alternative for enhancing segmentation reliability in noisy or low-contrast diagnostic imaging scenarios.

Qualitative visual comparisons in Fig. 3 further validate the efficacy of our method. It is evident that RBL produces significantly fewer red areas compared to other loss functions, demonstrating its superior capability in refining boundary details.

Beyond metric-based accuracy, RBL exhibits remarkable training stability. In our experiments, we observe that training with HD Loss and Dual HD Loss occasionally fails to converge (as indicated by the failed runs marked with "–" in Table I). Conversely, RBL exhibits demonstrates stable convergence behavior. The corresponding training curves on the CVC-ClinicDB dataset are illustrated in Fig. 4, underscoring its robustness as a reliable boundary-aware objective function.

As summarized in Table II, RBL introduces a manageable computational overhead compared to the baseline. Although its training time is slightly higher than the original Boundary Loss, this marginal increase is well-justified by the Improved precision it delivers. Notably, our RBL remains significantly more efficient than HD Loss and Dual HD Loss, which incur substantially higher costs regardless of the backbone. This favorable balance of speed and precision makes RBL a practical choice for real-world clinical applications.

In summary, the steady of RBL across diverse architectures and modalities underscores its potential as a robust strategy for medical image segmentation. By balancing high geometric accuracy and stable convergence with manageable computational overhead, it not only facilitates efficient model training but also offers a promising approach for enhancing generalization in varied clinical scenarios.

VI. Ablation Study

To evaluate the individual contribution of the proposed robust normalization strategy, we perform a comparative analysis across four diverse backbone architectures: nnU-Net, U-Mamba (bot), TransUNet, and SwinUNet. Three distinct configurations are examined: (1) No normalization, (2) Global Range normalization, and (3) RBL normalization. Experiments are implemented on the TN3K, BUSI, CVC-ClinicDB, and ISIC 2017 datasets, following the quantitative evaluation protocol established in Section IV-E.

The comparative results, summarized in Table III, reveal several key insights. First, the integration of RBL normalization consistently enhances boundary-related metrics (HD95 and ASD) across all tested architectures and modalities. Specifically, compared to the "No Norm" configuration, RBL significantly mitigates the gradient instability often encountered with raw distance maps, leading to more precise edge localization. Second, while the conventional global range normalization provides a moderate performance gain over the No Norm configuration by stabilizing the value range,

Authorized licensed use limited to: Shenzhen Technology University. Downloaded on October 26,2026 at 06:26:38 UTC from IEEE Xplore. Restrictions apply.

it remains sensitive to the skewed distribution of distance values. As a result, global range normalization typically achieves suboptimal performance compared to RBL, as it fails to effectively suppress extreme outliers that compress the informative boundary gradients. These findings underscore that the superior performance of our method stems not merely from normalization itself, but from the robust, quantile-based redistribution of distance information, which proves essential for handling the irregular morphologies typical in medical imaging.

VII. Conclusion

In this work, we proposed RBL, a simple yet effective boundary-aware loss function that significantly improves geometric accuracy in medical image segmentation. By introducing a robust normalization strategy based on quantiles for signed distance maps, RBL mitigates gradient instability caused by distributions dominated by outliers. Such distributions are commonly observed in low-contrast or noisy clinical images. Our method successfully preserves discriminative boundary information during this process.Extensive experiments across four datasets and multiple architectures show that RBL consistently improves boundary accuracy. Specifically, it reduces errors in HD95 and ASD while keeping the DSC at a high level. This confirms that RBL refines edges without losing overall segmentation quality. These results highlight the robustness and generalizability of the proposed method.

However, in the current version, the scheduling mechanism used to balance RBL with the base loss of the model still relies on empirical tuning for its weight coefficients that change over time. While our results confirm the effectiveness of this specific configuration, the optimal strategy to combine boundary-aware and region-based losses in a dynamic or adaptive way remains an open question. Future work should investigate principled approaches, such as weighting based on uncertainty or task aware scheduling, to automatically determine loss coefficients. Such methods could maximize segmentation fidelity across a broad range of medical imaging tasks and anatomical structures. Despite this limitation, RBL offers a plug-and-play solution for enhancing boundary precision in real-world clinical applications.

References

[1] C. Durante, G. Grani, L. Lamartina, S. Filetti, S. J. Mandel, and D. S. Cooper, "The diagnosis and management of thyroid nodules: a review," *Jama*, vol. 319, no. 9, pp. 914–924, 2018.

[2] P. Pimentel-Nunes, M. Dinis-Ribeiro, T. Ponchon, A. Repici, M. Vieth, A. De Ceglie, A. Amato, F. Berr, P. Bhandari, A. Bialek *et al.*, "Endoscopic submucosal dissection: European society of gastrointestinal endoscopy (esge) guideline," *Endoscopy*, vol. 47, no. 09, pp. 829–854, 2015.

[3] F. Nachbar, W. Stolz, T. Merkle, A. B. Cognetta, T. Vogt, M. Landthaler, P. Bilek, O. Braun-Falco, and G. Plewig, "The abcd rule of dermatoscopy: high prospective value in the diagnosis of doubtful melanocytic skin lesions," *Journal of the American Academy of Dermatology*, vol. 30, no. 4, pp. 551–559, 1994.

[4] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[5] F. Isensee, P. F. Jaeger, S. A. Kohl, J. Petersen, and K. H. Maier-Hein, "nnu-net: a self-configuring method for deep learning-based biomedical image segmentation," *Nature methods*, vol. 18, no. 2, pp. 203–211, 2021.

[6] J. Ma, F. Li, and B. Wang, "U-mamba: Enhancing long-range dependency for biomedical image segmentation," *arXiv preprint arXiv:2401.04722*, 2024.

[7] J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang, L. Lu, A. L. Yuille, and Y. Zhou, "Transunet: Transformers make strong encoders for medical image segmentation," *arXiv preprint arXiv:2102.04306*, 2021.

[8] H. Cao, Y. Wang, J. Chen, D. Jiang, X. Zhang, Q. Tian, and M. Wang, "Swin-unet: Unet-like pure transformer for medical image segmentation," in *European conference on computer vision*. Springer, 2022, pp. 205–218.

[9] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *2016 fourth international conference on 3D vision (3DV)*. Ieee, 2016, pp. 565–571.

[10] A. Mao, M. Mohri, and Y. Zhong, "Cross-entropy loss functions: Theoretical analysis and applications," in *International conference on Machine learning*. pmlr, 2023, pp. 23 803–23 828.

[11] X. Liu, J. Tian, S. Huang, and W. Shen, "Enhancing medical image segmentation via complementary cnn-transformer fusion and boundary perception," *Frontiers in Computer Science*, vol. 7, p. 1677905, 2025.

[12] Y. Xue, H. Tang, Z. Qiao, G. Gong, Y. Yin, Z. Qian, C. Huang, W. Fan, and X. Huang, "Shape-aware organ segmentation by predicting signed distance maps," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 34, no. 07, 2020, pp. 12 565–12 572.

[13] J. Ma, Z. Wei, Y. Zhang, Y. Wang, R. Lv, C. Zhu, C. Gaoxiang, J. Liu, C. Peng, L. Wang *et al.*, "How distance transform maps boost segmentation cnns: an empirical study," in *Medical Imaging with Deep Learning*. PMLR, 2020, pp. 479–492.

[14] Q. Lin, X. Chen, C. Chen, and J. M. Garibaldi, "Boundary-wise loss for medical image segmentation based on fuzzy rough sets," *Information Sciences*, vol. 661, p. 120183, 2024.

[15] M. B. Erden, S. Cansiz, O. Caki, H. Khattak, D. Etiz, M. C. Yakar, K. Duruer, B. Barut, and C. Gunduz-Demir, "Fourierloss: Shape-aware loss function with fourier descriptors," *Neurocomputing*, vol. 638, p. 130155, 2025.

[16] J. Ma, J. Chen, M. Ng, R. Huang, Y. Li, C. Li, X. Yang, and A. L. Martel, "Loss odyssey in medical image segmentation," *Medical image analysis*, vol. 71, p. 102035, 2021.

[17] A. Celaya, B. Riviere, and D. Fuentes, "A generalized surface loss for reducing the hausdorff distance in medical imaging segmentation," *arXiv preprint arXiv:2302.03868*, 2023.

[18] S. M. Hosseini, "topk dice loss for medical image segmentation." in *BMVC*, 2024.

[19] F. Xu, F. Yang, X. Li, and X. Zhang, "A unified loss for handling inter-class and intra-class imbalance in medical image segmentation," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 39, no. 8, 2025, pp. 8842–8850.

[20] D. Karimi and S. E. Salcudean, "Reducing the hausdorff distance in medical image segmentation with convolutional neural networks," *IEEE Transactions on medical imaging*, vol. 39, no. 2, pp. 499–513, 2019.

[21] H. Kervadec, J. Bouchtiba, C. Desrosiers, E. Granger, J. Dolz, and I. B. Ayed, "Boundary loss for highly unbalanced segmentation," in *International conference on medical imaging with deep learning*. PMLR, 2019, pp. 285–296.

[22] W. Al-Dhabyani, M. Gomaa, H. Khaled, and A. Fahmy, "Dataset of breast ultrasound images," *Data in brief*, vol. 28, p. 104863, 2020.

[23] H. Gong, G. Chen, R. Wang, X. Xie, M. Mao, Y. Yu, F. Chen, and G. Li, "Multi-task learning for thyroid nodule segmentation with thyroid region prior," in *2021 IEEE 18th international symposium on biomedical imaging (ISBI)*. IEEE, 2021, pp. 257–261.

[24] J. Bernal, F. J. Sánchez, G. Fernández-Esparrach, D. Gil, C. Rodríguez, and F. Vilariño, "Wm-dova maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians," *Computerized medical imaging and graphics*, vol. 43, pp. 99–111, 2015.

[25] Y. Li and L. Shen, "Skin lesion analysis towards melanoma detection using deep learning network," *Sensors*, vol. 18, no. 2, p. 556, 2018.

Authorized licensed use limited to: Shenzhen Technology University. Downloaded on October 26,2026 at 06:26:38 UTC from IEEE Xplore. Restrictions apply.