Enhancing Boundary Segmentation for Topological Accuracy with Skeleton-based Methods

Chuni Liu$^{1,2}$, Boyuan Ma$^{2,3,4,*}$, Xiaojuan Ban$^{2,3,4,5,*}$, Yujie Xie$^{2,3}$, Hao Wang$^{2,6}$, Weihua Xue$^{7}$, Jingchao Ma$^{1}$ and Ke Xu$^{1,*}$

$^{1}$Collaborative Innovation Center of Steel Technology, University of Science and Technology Beijing
$^{2}$Beijing Advanced Innovation Center for Materials Genome Engineering, University of Science and Technology Beijing
$^{3}$School of Intelligence Science and Technology, University of Science and Technology Beijing
$^{4}$Shunde Innovation School, University of Science and Technology Beijing
$^{5}$Institute of Materials Intelligent Technology, Liaoning Academy of Materials
$^{6}$Institute for Advanced Materials and Technology, University of Science and Technology Beijing
$^{7}$School of Materials Science and Technology, Liaoning Technical University
c Chuniliu@xs.ustb.edu.cn

Abstract

Topological consistency plays a crucial role in the task of boundary segmentation for reticular images, such as cell membrane segmentation in neuron electron microscopic images, grain boundary segmentation in material microscopic images and road segmentation in aerial images. In these fields, topological changes in segmentation results have a serious impact on the downstream tasks, which can even exceed the misalignment of the boundary itself. To enhance the topology accuracy in segmentation results, we propose the Skea-Topo Aware loss, which is a novel loss function that takes into account the shape of each object and topological significance of the pixels. It consists of two components. First, a skeleton-aware weighted loss improves the segmentation accuracy by better modeling the object geometry with skeletons. Second, a boundary rectified term effectively identifies and emphasizes topological critical pixels in the prediction errors using both foreground and background skeletons in the ground truth and predictions. Experiments prove that our method improves topological consistency by up to 7 points in VI compared to 13 state-of-art methods, based on objective and subjective assessments across three different boundary segmentation datasets. The code is available at https://github.com/clovermini/Skea_topo.

1 Introduction

Although deep learning algorithms for boundary segmentation show high accuracy in pixel-level metrics, achieving topologically consistent predictions that align with the ground truth remains a significant challenge. This is particularly evident in the boundary segmentation of reticular images, such as neuron images [Plaza et al., 2014], material microscopic images [Liu et al., 2022] and aerial images [Mnih, 2013]. Topologically consistent segmentation demands the preservation of key attributes such as the connectivity, number of holes, and layout of segments. These attributes are crucial for the accuracy of subsequent analyses such as studying the structure and connectivity of neuronal circuits, quantitatively characterizing the microstructure, and navigation planning [Batra et al., 2019]. In boundary segmentation, errors that cause topological changes can be divided into four types, as illustrated in Fig. 1. When viewed in terms of objects such as grains and cells, these errors manifest as the misclassifica-
tion of critical pixels, which results in unusual appearances, unusual disappearances, and object merging or splitting.

Even a small number of misclassified pixels in the prediction result can lead to these errors with significant topological changes. Thus, standard pixel-based loss functions such as the cross-entropy (CE) loss and Dice loss exhibit inadequate topological correctness. These methods evaluate each pixel equally regarding the classification precision, so the overall loss can reach a low value even when the topological structure is incorrect. Some approaches such as [Liu et al., 2022; Valverde and Tohka, 2023] assess the misclassification severity based on the location of each pixel and adaptively weight each pixel within the objects (background) according to their distance from the boundary (foreground), as dictated by a distance function. This technique enhances the accuracy and topological consistency of the segmentation. Nevertheless, existing distance-based weighting methods struggle to accurately represent irregular object shapes and may fail to appropriately allocate greater weight to error-prone areas, such as narrow regions in irregular objects. In addition, they are unaware of whether the segmentation results have topological errors. Methods such as [Hu et al., 2019; Hu, 2022] focus on identifying topological critical pixels during training based on network predictions. These are specific pixels in binary images whose value reversal would cause topological changes. These methods place a greater emphasis on the accurate classification of topological critical pixels, which significantly enhances the topological consistency of the segmentation results. However, they currently suffer from noise, a lack of optimal penalty strategies, and time-consuming computations.

In this work, we propose a novel loss function called Skeaw Topo Aware to address the aforementioned challenges. Our approach incorporates two key components: a skeleton-aware weighted loss (Skeaw) and a boundary rectified term (BoRT).

Skeaw introduces an innovative weighting approach that applies to both foreground and background pixels, considering the importance of each pixel based on the location and overall category quantity ratio. For the foreground, the weight is calculated based on the overall boundary thickness: a thinner boundary is assigned a larger weight, since it is more fragile and prone to fracture. For the background, Skeaw enhances the existing distance-weighting function by employing the object’s skeleton to refine the modeling of its shape, which ensures that the weighting results adequately account for the shape features of irregular objects. The proposed skeleton-based approach leverages the benefits of existing distance-weighted methods while overcoming their limitations for irregular objects.

The core motivation of BoRT is to more rapidly identify the topological critical pixels in errors and apply superior penalty methods to steer the model towards generating topologically consistent results. An in-depth analysis of the error regions that cause topological changes has led to the development of a simpler and more efficient method, which serves as an alternative to the existing time-consuming identification techniques. This method identifies topological critical pixels by leveraging the trait that they consistently coincide with the foreground or background skeleton in the ground truth and predictions. Conversely, non-topological critical pixels typically do not intersect with these skeletons, except for errors distant from the boundaries, which also necessitate increased penalties. By using the foreground and background skeletons, BoRT can accurately identify connected regions, where topological critical pixels are located, through a process of high computational efficiency. Based on this identification, a novel penalty term, which assigns higher penalties to critical pixels and reduces the penalty strength for non-critical pixels, is devised using likelihood maps to effectively improve the topological consistency of the segmentation result.

In general, our contributions are highlighted as follows:

- We propose a skeleton-aware weighted loss, which better utilizes the geometric information of objects to establish a weighted map. This approach is more conducive to preserving the geometric and topological properties of irregular objects during the learning process.
- We propose a novel boundary rectified term to efficiently and effectively identify and emphasize topological critical pixels that affect the topology during training. This term can guide the network to effectively preserve the topology of the boundary segmentation results.
- We conduct extensive experiments for three different boundary segmentation tasks: road segmentation, cell membrane segmentation, and grain boundary segmentation. Our method performs better objective and subjective assessments than the state-of-the-art (SOTA) methods. In addition, the boundary rectified term can be used as a plug-in for other classical loss functions to further improve the performance.

2 Related Work

In recent years, deep learning-based boundary segmentation methods have undergone rapid development. By introducing the fully convolutional layer, [Long et al., 2015] enables pixel-level image classification and establishes the foundation for a deep learning image segmentation model. Many methods have since optimized the model to improve the performance and accuracy from various angles. The UNet model, which was introduced by [Ronneberger et al., 2015], presents an elegant symmetrical encoding and decoding structure. It incorporates residual connections in the middle layers, which effectively preserve intricate details from the original image. UNet and its variants [Zhou et al., 2019; Isensee et al., 2021; Saadati et al., 2023; Roy et al., 2023] have gained widespread popularity due to their excellent and robust performance in boundary segmentation tasks. Our work takes UNet as a baseline to assess different loss functions, and our approach is adaptable to various network architectures.

For deep learning models that are optimized using a gradient descent [LeCun et al., 1988], a common approach is to adjust the learning direction of the network by designing a target-based loss function. To preserve the topology, various loss functions have been proposed to directly or indirectly increase the penalty when topological errors occur.

Weighted losses. Certain losses prioritize weighting areas that are prone to topological errors, which emphasizes the
segmentation correctness on boundaries. For example, [Long 
et al., 2015] introduced a class balance cross-entropy loss (BCE) that applied heavier weights to boundary pixels based on the number of categories. The Boundary loss [Kervadec et al., 2019] calculates the loss based on boundary errors only. GraphCuts loss [Zheng et al., 2021] add boundary penalty term based on the graph cuts theory. Dice loss [Milletari et al., 2016], Tversky loss [Salehi et al., 2017], Lovasz loss [Berman et al., 2018], and margin loss [Yu et al., 2022] collectively treat pixels of the same class and emphasize the overall boundary segmentation accuracy. These methods overlook the varying topological importance of pixels located at different positions. UNet [Ronneberger et al., 2015] introduces a weighted cross-entropy loss (WCE), which considers the distance between boundary pixels and nearby objects. However, for evenly spaced objects, it is equal to the BCE. Some losses weight the pixels in each object based on a distance transform function, such as the error-penalizing distance weighted loss (EPDW) [Caliva et al., 2019], adaptive boundary weighted loss (ABW) [Liu et al., 2022], and region-wise loss (RW) [Valverde and Tohka, 2023]. However, their weighting methods are not suitable for irregular objects, which have narrow or curved regions that are difficult to classify. These regions have lower weights because their pixels are close to the boundary. Our method solves this problem by normalizing the weights using the skeleton.

**Topology-aware losses.** Many methods more directly improve the model’s perception of the topological structure of images to enhance the topological consistency of segmentation results. Topoaware loss [Mosinska et al., 2018] and RMI loss [Zhao et al., 2019] model the output structure and enforce its consistency, but this structure may be unrelated to the topology. cIDice loss [Shit et al., 2021] and DMT loss [Hu et al., 2021] define the skeleton of the boundary as the topological critical structures and enforce these locations to be correctly segmented. However, these pixels may not be the ones that actually cause topology errors. [Turaga et al., 2009; Funke et al., 2018] adopt the idea of the rand index, find errors in the affinity graph that caused split and merge errors by calculating the maximin path among pixels and imposed strong penalties. Based on this method, [Oner et al., 2021] focuses on the connectivity of background regions in the predicted distance map and penalizes the unwanted interruptions and false connections. However, the computational complexity of comparing pixel pairs is very high, and focusing on a single pixel with the smallest value can lead to unstable training. Many methods use persistent homology-based methods to describe and compare the topological structures of images. [Hu et al., 2019; Oner et al., 2023] compare the persistence diagrams to identify critical pixels in the likelihood map, but they can be easily affected by noise. [Stucki et al., 2023] uses induced matchings among persistence barcodes to precisely formalize the spatial correspondences between topological structures of two images. However, the computational complexity of persistent homology methods is cubic in relation to image size, making them time-consuming for large datasets. [Jain et al., 2010] proposes a warping error method, which can accurately determine whether a pixel can cause a topological change based on the digital topological knowledge. Warp loss [Hu, 2022] accelerates it through the proposed efficient Distance-Ordered Homotopy Warping method, which identifies critical pixels through homotopic warping pixels on the discrepancies between predicted results and ground truth in order of the distance. However, it requires a pixel-wise operation, and its penalty method is not optimal. In our work, we propose a much more efficient approach to accurately identify topological critical pixels and design a better penalty term.

### 3 Method

The proposed Skea-Topo Aware loss consists of a Skeleton Aware Weighted loss and a Boundary Rectified Term (as shown in Fig. 2 (a)). The former precomputes the weight map before training. The latter identifies topological critical pixels during training. These two components complement each other and collaborate in the overall process. Therefore, the total loss function is defined as follow:

$$L_{\text{total}} = L_{\text{skeaw}} + \lambda L_{\text{bortjomo}}$$

where $L_{\text{skeaw}}$ is the Skeleton Aware Weighted loss, $L_{\text{bortjomo}}$ is the Boundary Rectified Term, and the weight $\lambda$ is utilized to balance the two terms.

#### 3.1 Skeleton Aware Weighted Loss

By using a combination of a pixel-wise weight map and the cross-entropy loss, the attention of the network towards specific pixels can be regulated. For the boundary segmentation task in this work, each pixel is classified as an object (background, class 0) or a boundary (foreground, class 1). The Skeleton Aware Weighted loss is formulated as follows:

$$L_{\text{skeaw}} = -\sum_{x \in \Omega} \begin{cases} w_1^s(x) \log(p_1(x)), & \text{if } w_0^b(x) < w_1^s(x) \cdot m_d(x) \\ w_0^b(x) \log(p_0(x)), & \text{if } w_0^b(x) \geq w_1^s(x) \cdot m_d(x) \end{cases}$$

where $p_0(x)$ and $p_1(x)$ are the probabilities obtained after applying the softmax activation function. $w_1^s(x)$ and $w_0^b(x)$ are the weighted values assigned to each pixel $x \in \Omega, \Omega' \subset \mathbb{R}^{2,3}$ in the foreground and background, respectively. $m_d(x)$ is the dilation mask of the ground truth, the actual foreground and background are determined by comparing the weight of the background with that of the dilated foreground. This approach increases the tolerance for slight offsets between predicted and ground truth boundaries, which can improve the segmentation accuracy. The degree of dilation is controlled by the hyperparameter $d_{\text{iter}}$.

The design of our loss is based on two principles. First, for the boundary, higher weights should be assigned to pixels in narrower boundaries, since they have a higher risk of fracture. Second, for the object, higher weights are assigned to misclassified pixels far away from the boundary, since these near it can be difficult to distinguish, often due to boundary ambiguity and manual annotation misalignment. Existing weighting methods use the distance transform function to assign weights to individual objects. For the stability of training, these methods normalize the distance value using the
maximum distance, so relatively low weights are assigned to narrow areas of irregular objects, which are prone to boundary closure errors. To overcome this limitation, our method introduces the object skeleton for distance calculation and weight normalization. The skeleton of an object is extracted through the morphological skeletonization function provided by scikit-image [Van der Walt et al., 2014]. Through the object skeleton, a uniform weight distribution can be achieved in narrow areas and other areas, as shown in Fig. 2 (c). The detailed definitions of \( w_1(x) \) and \( w_0(x) \) are as follows:

\[
\begin{align*}
    w_1(x) &= w_{1e}^b(x) + w_0 \times \left( \frac{2d_{1}^{\max} - d_1(x) - d_2(x)}{2d_{1}^{\max}} + 1 \right) \quad (3) \\
    w_0(x) &= w_{0e}^b(x) + w_0 \times \left( \frac{d_{nsp}(x)}{d_{0}^{nsp}} \right) \quad (4)
\end{align*}
\]

where \( d_1(x) \) and \( d_2(x) \) are the euclidean distance from boundary pixel \( x \) to the closest and next closest objects, respectively. \( d_{1}^{\max} \) is the maximum value among all \( d_1(x) \). \( d_{nsp}(x) \) is the distance from pixel \( x \) within an object to the nearest boundary pixel, and \( d_{0}^{nsp} \) is the distance from the above nearest boundary pixel to the closest skeleton point of pixel \( x \), as shown in Fig. 2 (c). The weights \( w_{1e}^b(x) \) and \( w_{0e}^b(x) \) serve as class-balance weights and are equivalent to the inverse proportion of each category. \( w_0 \) is a hyperparameter to assign the weight of the weight map, which is set to 10 by default in our experiments.

### 3.2 Boundary Rectified Term

Before introducing the algorithm, several concepts are defined for better understanding. Positive (\( p \)) refers to the pixels predicted as the boundary; negative (\( n \)) refers to the pixels predicted as objects; true (\( t \)) signifies matching predicted and real classes; false (\( f \)) signifies mismatched predicted and real classes. True positives (\( tp \)) and true negatives (\( tn \)) represent correct predictions; false positives (\( fp \)) and false negatives (\( fn \)) indicate incorrect predictions. The errors in this work are the combination of \( fp \) and \( fn \). Topological critical pixels are equal to the sum of topological false positives (\( tfp \)) and topological false negatives (\( tfn \)) and abbreviated as true errors (\( tf \)), which mean errors that lead to topological changes. Then, \( tfp \) and \( tfn \) are non-topological critical errors, which do not affect the topology, and abbreviated as false errors (\( ff \)).

Fig. 2 (b) shows the step-by-step legend of the calculation process of \( tfn \). Considering the four situations that lead to topological changes, the topological critical pixels that correspond to boundary incorrect closure and object disappearance are only found in \( fp \), whereas those linked to boundary incorrect fracture and object appearance only occur in \( fn \). Taking these two situations in \( fn \) as examples for analysis, when a new object appears in the prediction, its skeleton must intersect with \( fn \), so the corresponding connected region of critical pixels can be found by calculating the intersection of the background skeleton of prediction and \( fn \). The background skeleton is derived from the sum of the skeletons of each object. The background skeleton can also be used to identify most incorrect fracture regions. Typically, a fractured
area becomes a narrow passageway that connects two merged objects, and the background skeleton invariably crosses this bridge area. However, in certain extreme cases such as the one highlighted in the yellow box in Fig. 2 (b), the background skeleton fails to detect topological critical pixels. In these instances, the non-overlapping area between the foreground skeleton of the ground truth and the prediction accurately pinpoint the fractured region. This foreground skeleton method does not effectively identify new object regions, as demonstrated in the yellow circle in Fig. 2 (b). When these two approaches are used together, they complement each other and collectively cover all potential scenarios. This synergy enables the precise identification of all topological critical pixels. For disappearing objects and incorrect closed boundaries, the same logic applies when the labels are treated as predictions and vice versa.

Skeletons are used because they can effectively exclude most of the $ff$ near the boundary but not areas that significantly alter the shape or are located far from the boundary. Such errors should also be assigned a heavier penalty to maintain the accuracy of the results. Instead of individually examining each pixel to determine whether flipping that pixel will change the topology as Warp[Hu, 2022], our approach identifies all critical pixels using a few image-level operation steps through the skeletons. The most time-intensive step of this process is computing the foreground and background skeletons, each with a time complexity of $O(n)$, where $n$ represents the number of image pixels.

Our method directly multiplies the probability map with the identified critical pixels to define the penalty term. The key goal is to enable $ffn$ and $ffp$ to be classified into opposing categories while ensuring the accurate classification of $tn$, $tfn$, and $tfp$. Essentially, this step adjusts the actual training target based on the detected $tf$ and $ff$ during training. Therefore, this method is named the boundary rectified term. The loss function is calculated using the rectified likelihood map $p_{l(x)}^{fix}(x)$ (where $l \rightarrow \{0, 1\}$), and $l(x)$ is the class of pixels), which is derived from the following formula:

$$L_{bort_{topo}} = \sum_{x \in \Omega} p_{l(x)}^{fix}(x)$$

(5)

$$p_{0}^{fix}(x) = tn \times p_{1}(x) + tfp \times p_{1}(x) + ffp \times p_{0}(x)$$

(6)

$$p_{1}^{fix}(x) = tp \times p_{0}(x) + tfn \times p_{0}(x) + ffn \times p_{1}(x)$$

(7)

where $p_{l(x)}$ is the likelihood map. While not explicitly stated in the formula, $tfp$ and $tfn$ can be properly weighted according to the data situation to obtain better results. Our formula is similar to Poly1[Leng et al., 2022]. Poly1 adjusts the importance of the first polynomial coefficients of cross-entropy loss, which results in higher segmentation accuracy. A comparison experiment between Poly1 and BoRT will be conducted to further validate the effectiveness of identifying topological critical pixels.

4 Experiments

Extensive experiments were conducted to validate the effectiveness of our proposed method in preserving the topology. All experiments were performed using an NVIDIA GeForce RTX 3090 GPU (24GB Memory) and an Intel(R) Xeon(R) Silver 4210R CPU @ 2.40 GHz.

4.1 Experiment Setting

Datasets

Our method was evaluated using three publicly available image segmentation datasets from different domains. The first dataset, SNEMI3D [Arganda-Carreras et al., 2013] in biology, consists of electron microscopy (EM) images to segment mouse cortex neurites in 3D. It has a dataset size of $100 \times 1024 \times 1024$, and its 2D slices were used here for the 2D boundary segmentation experiments. The second dataset, Pure Iron Grain(IRON)[Liu et al., 2022], focuses on pure iron grain segmentation in material science. It comprises 296 microscopic slices with an image size of $1024 \times 1024$. The third dataset, Massachusetts Roads Dataset (MASS. ROAD)[Mnih, 2013] from remote sensing, contains 1171 aerial images of Massachusetts, each of which is $1500 \times 1500$ pixels. To assess the model robustness, a three-fold cross-validation approach was used for the SNEMI3D and IRON datasets. For the MASS. ROAD Dataset, validation and testing were conducted using its official sets.

Baselines and Implementation Details

In our experiments, we compared our method with 13 SOTA loss functions (Table 1). All methods were trained based on the standard PyTorch implementation of the UNet model. To ensure fairness, a consistent configuration was maintained. In the preprocessing stage, all images were normalized based on the mean and standard deviation(std) calculated on each dataset. The data augmentation techniques of random crop, rotation and flip were applied. The input size was set to $512 \times 512$. During training, the Adam optimizer with default parameters was used, and the StepLR scheduler with a step size of 10 and a decay rate of 0.8 was used to adjust the learning rate. The initial learning rate was set to 1e-4. Each model was trained for 50 epochs with a batch size of 10, and we obtained the best parameter based on early stopping. These parameter values were selected through a grid search. All SOTA loss functions were implemented using its official PyTorch implementation. Poly1 [Leng et al., 2022] based on the cross-entropy loss was used. Hyperparameter $\alpha$ for cIDice [Shit et al., 2021] was set to 0.1 through a grid search. For datasets with evenly spaced objects such as IRON and MASS. ROAD, WCE [Ronneberger et al., 2015] is equivalent to BCE [Long et al., 2015], since all foreground pixels have identical weights.

Evaluation Metrics

Five metrics were used in the experiment to evaluate the performance of each method. These metrics include a pixel-wise metric (Dice) and topology-related metrics: Variation of Information (VI) [Nunez-Iglesias et al., 2013], Adjusted Rand Index (ARI) [Vinh et al., 2009], mean average precision (mAP) [Lin et al., 2014], and Betti Error (Betti). VI is given prominence in our evaluation because it is sensitive to the presence of topological changes and the extent of their impact. To ensure fair evaluation, post-processing was applied to the segmentation results and ground truth. It includes
Three steps: boundary dilation with a square filter of size 3 to remove small holes, skeletonization, and a second round of dilation using the same filter.

4.2 Qualitative and Quantitative Results

Table 1 displays the quantitative results of our loss function compared to 13 other SOTA losses. Our method delivers top-tier performance across most metrics: it ranks first in most cases and a close second in some cases. Other methods exhibit their own strengths and weaknesses on different metrics, whereas our approach excels across all evaluated metrics, which showcases its well-rounded effectiveness. Compared to the existing methods, our approach achieves a maximum improvement of 7 points in VI. Significantly, no other method has simultaneously shown superior performance across all three datasets as ours does, which underscores the exceptional effectiveness and robustness of our approach. The subscript data with ± represent the std of the results of three-fold experiments. The relatively low std of our method highlights its stability. All reported metrics of our method are based on the best overall performance in the ablation studies. Figs. 3, 4, and 5 qualitatively compare the boundary segmentation results between our method and other methods across three datasets. These figures demonstrate that our method is more effective than other methods in eliminating topological errors when compared to the ground truth (GT).

4.3 Ablation Studies

This section reports only the ablation experiment results on the SNEMI3D dataset. The loss weight λ was set to 1 by default, which already significantly improves performance.

Choice of Hyperparameters

There are two critical hyperparameters involved in the proposed loss function. The first one is the dilation iteration (dilation) in Skaw, which is designed to broaden the coverage of the foreground. A value of 1 in this context indicates a dilation operation using a square filter with a size of 3. The second hyperparameter determines the epoch at which we initiate performance across most metrics. These figures demonstrate that our method is more effective than other methods in eliminating topological errors when compared to the ground truth (GT).

Figure 3: Qualitative results of different losses on the SNEMI3D dataset. The red arrows and blue rectangles indicate topological false positive and false negative errors, respectively.

Figure 4: Qualitative results of different losses on the IRON dataset.
the best performance and a relatively low standard deviation, which indicates stable training. The findings suggest that dilating the ground truth can improve the segmentation accuracy. The trends in Fig. 6 (a) also reveal that the effectiveness of BoRT is closely tied to the timing of its inclusion, notably, setting the parameter to 10 leads to a noticeable decrease in performance. In subsequent experiments, step_num = 20 is used as the default setting.

Impact of the Loss Terms

Ablation experiments on each component of our proposed loss were performed to verify their necessity. Specifically, the components of the penalty term in BoRT were evaluated, w/o \( ff \) refers to removing the \( f f p \) and \( f f n \) terms in \( L_{bort\_topo} \), and w/o \( tt \) indicates removing the \( t n \) and \( t p \) terms. The results in Table 2 demonstrate the necessity of all penalty components. Furthermore, similar ablation studies were conducted on the other two datasets. The IRON and SNEMI3D datasets have generally consistent conclusions, but the standard form of the penalty term poorly performs on the MASS. ROAD dataset, possibly due to the extreme data imbalance and setting the parameter to 10 leads to a noticeable decrease in performance. In subsequent experiments, \( step\_num = 20 \) is used as the default setting.

<table>
<thead>
<tr>
<th>Method</th>
<th>VI↓</th>
<th>mAP↑</th>
<th>ARI↑</th>
<th>Dice↑</th>
<th>Betti↓</th>
</tr>
</thead>
<tbody>
<tr>
<td>w/o ( ff )</td>
<td>39.52±4.38</td>
<td>45.54±2.84</td>
<td>63.59±2.32</td>
<td>75.69±0.99</td>
<td>17.59±7.31</td>
</tr>
<tr>
<td>w/o ( tt )</td>
<td>39.67±5.76</td>
<td>44.65±1.70</td>
<td>63.80±0.74</td>
<td>75.88±0.57</td>
<td>20.72±3.50</td>
</tr>
<tr>
<td>w/o ( tt &amp; ff )</td>
<td>48.04±2.55</td>
<td>37.36±0.27</td>
<td>50.07±0.56</td>
<td>74.54±1.11</td>
<td>61.25±7.88</td>
</tr>
<tr>
<td>BoRT</td>
<td>36.59±1.94</td>
<td>47.49±0.88</td>
<td>65.49±1.21</td>
<td>76.43±0.26</td>
<td>12.72±2.68</td>
</tr>
</tbody>
</table>

Table 2: Ablation study for loss components. The bold values indicate the best performance. BoRT means Skea-Topo here.

The Efficiency and Effectiveness of BoRT

To further demonstrate the efficiency and effectiveness of BoRT, thorough ablation experiments were compared to the advanced critical pixels detection-oriented Warp method. Specifically, we compared Warp, WaRT, and BoRT based on the Dice loss (identical to the Warp loss defined in [Hu, 2022])s under identical conditions). WaRT employs the Warp approach to identify topological critical pixels and integrates the penalty term of BoRT into its loss function. Fig. 6 (c) shows that BoRT delivers the best performance, and WaRT surpasses Warp in various metrics. Notably, the processing time of Warp is almost double that of BoRT. These results convincingly demonstrate that the method of BoRT to identify topological critical pixels is both efficient and effective. Furthermore, the penalty term introduced in BoRT proves more effective in promoting the network to preserve a topology. The comparison with Poly1, as shown in Fig. 6 (b), further confirms the contribution of BoRT.

Combining BoRT With Other Losses

Our proposed BoRT serves as a lightweight plug-and-play loss function aimed at enhancing the effectiveness of various segmentation loss functions. To showcase its versatility, we integrated BoRT into five distinct loss functions: CE, Dice, BCE, ABW, and Skea. The results illustrated in Fig. 6 demonstrate that incorporating BoRT leads to varying degrees of optimization, with improvements in the VI index ranging from 0.66 (Skeaw) to 27.5 (CE).

5 Conclusion

In this study, a Skea-Topo Aware loss was developed to guide the boundary segmentation network towards generating topologically consistent results. It includes a skeleton-aware weighted loss that uses object skeletons to model geometric information accurately, and a boundary rectification term that penalizes topologically critical pixels. A method based on foreground and background skeletons was proposed to identify these critical pixels efficiently. Our method was rigorously tested through comparative and ablation experiments on three datasets, showing an improvement of up to 7 points in the VI index, proving its effectiveness.

Limitations. While extensively studied on 2D images, extending it to multi-class and 3D segmentation is theoretically feasible but requires further validation in future work.
Acknowledgments
This work was supported by the National Key R & D Program of China under Grant 2022ZD0118001, the National Natural Science Foundation of China under Grant U22A2022 and 62106019, Scientific and Technological Innovation Foundation of Shunde Graduate School, USTB under Grant BK22BF010, and Fundamental Research Funds for the Central Universities of China under Grant 00007467, and China Postdoctoral Science Foundation under Grant 2021M700383. The computing work is supported by USTB MatCom of Beijing Advanced Innovation Center for Materials Genome Engineering.

Contribution Statement
Chuni Liu and Boyuan Ma contributed equally to this work.

References


[Mosinska et al., 2018] Agata Mosinska, Pablo Marquez-Neila, Mateusz Kozinski, and Pascal Fua. Beyond the...


