REASSESSING UTILITY OF TOPOLOGY BASED LOSSES FOR IMAGE SEGMENTATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Image segmentation is an important and widely performed task in computer vision. Accomplishing effective image segmentation in diverse settings, often requires custom model architectures and loss functions. A set of models that specialize in segmenting thin tubular structures are topology preservation based loss functions. These models often utilize a pixel skeletonization process claimed to generate more precise segmentation masks of thin tubes and better capture the structures other models often missed.

One such model, Skeleton Recall Loss (SRL) proposed by Kirchhoff et al (Kirchhoff et al., 2024), was stated to produce state-of-the-art results on benchmark tubular datasets. In this work, we tested the validity of the SRL loss by using two approaches: empirically and theoretically. Upon comparing the performance of the proposed method on some of the tubular datasets (used in the original work, along with some additional datasets), we found that the performance of SRL based segmentation models did not exceed traditional baseline models. We then go on to examine and provide a theoretical explanation as to why losses based on topology based enhancements (including the SRL) fail to fulfill their objective.

1 PROBLEM STATEMENT AND THE PROPOSED SOLUTION

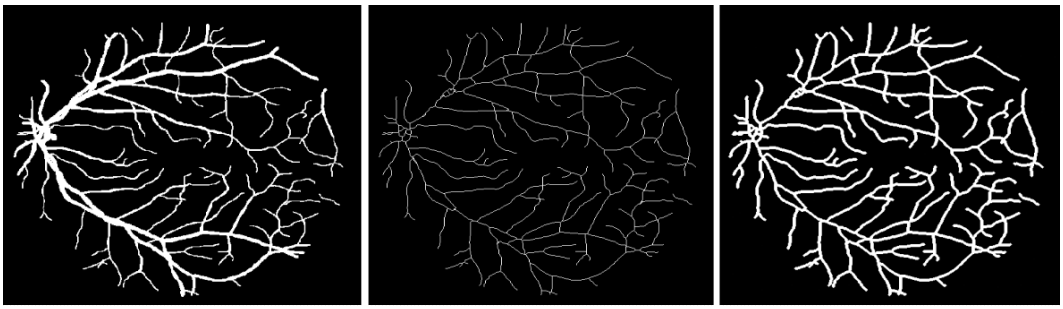
Image segmentation models help identify and accentuate structures of interest within an image. These models need to be versatile since they have to accurately delineate objects of diverse shapes, sizes and texture. Thin, tubular and curvilinear structures are particularly challenging for these segmentation models due to a scarce number of pixels corresponding to the region of interest. Common examples of such tasks include (but are not limited to): roads in satellite imagery, blood vessels, capillaries & canals in medical images and histology images. Multiple modifications (Shit et al., 2021; Cheng et al., 2021; Menten et al., 2023) have been proposed to perform segmentation of thin tubular structures efficiently.

One category of such methods are the specialized topology preservation based loss functions that utilize a skeletonization process whereby they reduce a shape in a binary image to its one-pixel-wide connected center-line that preserves its overall topology. These processes are meant to reduce the foreground to a thin network of pixels which is representative of the topology of original foreground. Different segmentation methodologies utilize different types of skeletonization processes (Jin & Kim, 2017; Abu-Ain et al., 2013; Latecki et al., 2007), often followed by the employment of skeletons thus formed in a specialized loss function. One such proposed method was the Skeleton Recall Loss (Kirchhoff et al., 2024), specialized for the segmentation of thin tubular and curvilinear structures.

In this work, we investigate the idea of using skeletonization for topology preservation by focusing on one such method presented as SRL (Kirchhoff et al., 2024) which claimed to have achieved state-of-the-art results for improving the segmentation of thin tubular structures by utilizing skeletons. nnU-Net(Isensee et al., 2018) based models were trained to perform segmentation of thin tubular structures, with and without the SRL loss, to analyze the impact of the SRL loss. Along with the empirical observations, we perform a theoretical analysis of the gradients of such loss functions and explain the shortcomings of the proposed methodology.

2 FORMULATION AND RESULTS FROM THE SKELETON RECALL LOSS

In the paper (Kirchhoff et al., 2024), a new binary image transformation technique is introduced as Tubed Skeletonization. The procedure first involves forming a skeleton of the ground truth mask using any of the standard skeletonization techniques (Abu-Ain et al., 2013; Jin & Kim, 2017; Latecki et al., 2007). Then, a tubed skeleton of the ground truth mask is formed by dilating the single pixel thin skeleton.



(a) GT segmentation (b) Skeleton (c) Tubed Skeleton

Figure 1: Visual comparison of (b) normal skeleton formed using standard skeletonization algorithm and (c) the proposed tubed skeleton(formed by dilating (b) normal skeleton) used for Skeleton Recall Loss for (a) ground truth segmentation, originating from the DRIVE dataset (Hassan et al., 2015).

The difference between the original mask, a regular skeleton and tubed skeleton is shown in Figure 1. This tubed skeleton is employed as the transformed ground truth to calculate the SRL with the following function:

$$\mathscr{L}_{SRL} = -\frac{1}{|K|}\sum_{k\in K}\frac{\sum_{i\in\Omega} s_\theta^{ik}.y^{ik}}{\sum_{i\in\Omega} y^{ik}} \tag{1}$$

where Ω is the set of pixels in segmentation mask, K is the set of all classes present in the ground truth mask, s_θ^{ik} denotes the ith pixel of predicted mask for the kth class, and θ are the parameters that affect s^{ik}, and y^{ik} denotes the ith pixel of Tubed Skelton formed using the ground truth mask for the kth class.

This loss function is used along with the standard loss functions while training the model. The net loss is then defined as

$$\mathscr{L}_{net} = \mathscr{L}_{generic} + \alpha. \cdot \mathscr{L}_{SRL}$$

Here, α is a hyper parameter that can be tuned to alter the contribution of the specialized loss function towards the net loss.

To validate the effectiveness of SRL in predicting thin tubular structure, a wide of variety of datasets were utilized by the authors of the SRL paper (Kirchhoff et al., 2024). These include natural segmentation tasks like Roads(Mnih, 2013) and Cracks (Tomaszkiewicz & Owerko, 2023), along with medical datasets which include 2D datasets like DRIVE(Hassan et al., 2015) and 3D datasets like TopCow (Yang et al., 2024). Though the test metrics on models trained using SRL did not show significant difference from the vanilla models, the paper claimed to have achieved better tubular segmentation. The visual results presented the claim that SRL trained model is capable to predict the tubes which are missed by other models. The paper concluded that the presented algorithm is able to enhance the tubular segmentation at a very marginal computational cost. The skeleton used for this procedure does not necessarily have to be continuous, and hence their algorithm is able to adapt to all kinds of skeletonization techniques (Abu-Ain et al., 2013; Jin & Kim, 2017; Latecki et al., 2007).

3 EXPERIMENTAL SETUP

3.1 DATASET DESCRIPTION

We tried to reproduce the results of SRL on two of the datasets mentioned in the SRL paper (Kirchhoff et al., 2024), namely DRIVE (Hassan et al., 2015) and Cracks (Tomaszkiewicz & Owerko, 2023). DRIVE is a medical dataset focused on segmenting blood vessels from retinal fundus images. Cracks, on the other hand, is a dataset for the segmentation of thin cracks in concrete. Both datasets are 2D and involve binary classification. Appropriate train-test splits were made for each dataset. The data pre-processing was handled by no new U-Net (nnUNet) (Isensee et al., 2018) itself. Some additional tests were performed on non-tubular datasets, like BoMBR(Raina et al., 2024) and ACDC(Bernard et al., 2018), to study the effect of topology preservation based loss function on such datasets, as mentioned in appendix.

3.2 MODEL ARCHITECTURE AND TEST METRICS

To replicate the experiments, we employed the nnUNet architecture. Two separate models were trained for each dataset. Each model was trained on 5 folds. The first model was trained using standard Dice loss (Dice) and Categorical Cross Entropy loss (CCE). The other model utilized SRL with the standard loss functions. The models were trained on NVIDIA T1000 GPU with 8GB of VRAM. The code for this implementation was directly taken from **github.com/MIC-DKFZ/Skeleton-Recall**.

We included multiple metrics covering aspects of overlap and topology preservation to comprehensively evaluate the performance of the different segmentation models. Among these, centerlineDice (clDice) (Shit et al., 2021), that evaluates the center line of the structures while simultaneously considering overlap, has proven to be a particularly valuable metric for assessing topology preservation. For overlap-based evaluation, we utilized the widely used Dice Similarity Coefficient (DSC) and Jaccard Similarity Index (JSI) metrics. We used False Negative Rate (FNR) to gauge the portions of ground truth which are missed by the model, and hence it gives information about maintenance of connectivity. The False Positive Rate (FPR) was used to ensure the model wasn't over predicting positives merely to achieve a low FNR.

4 RESULTS AND DISCUSSION

The metrics in Table 1 present competing results between vanilla model and SRL loss based model on Cracks(Tomaszkiewicz & Owerko, 2023) and DRIVE(Hassan et al., 2015) datasets. The results in the table conclusively indicate that SRL fails to enhance performance on the given tubular datasets. Incase of non-tubular datasets, presented in appendix, SRL even seems to worsen the model performance significantly in case of BoMBR (Raina et al., 2024), while it shows negligible effects for ACDC (Bernard et al., 2018) dataset.

From the visual results for the given test instance, we can infer that both vanilla and SRL models were able to perform equally well in segmenting thick tubular structures. However, several thin tubular structures were overlooked by the SRL models, even though vanilla models could detect them, as shown in Figure 2.

Higher value of FPR in case of both datasets show that the model starts over predicting when SRL is utilized. Also it could not connect meaningful regions when needed as shown by the higher FNR value in DRIVE dataset. Thus, the SRL based models do not seem to outperform models based on traditional loss functions.

5 WHY IS SRL NOT WORKING AS EXPECTED?

To understand the effect of SRL on the model training, we propose to investigate the flow of gradients through the neural network due to SRL during back propagation. Understanding the gradients would be beneficial for interpreting their effects in modification of parameters. The net loss used to train nnUNet using SRL given by:

$$\mathscr{L} = \mathscr{L}_{DCE} + \mathscr{L}_{CE} + \mathscr{L}_{SRL}$$

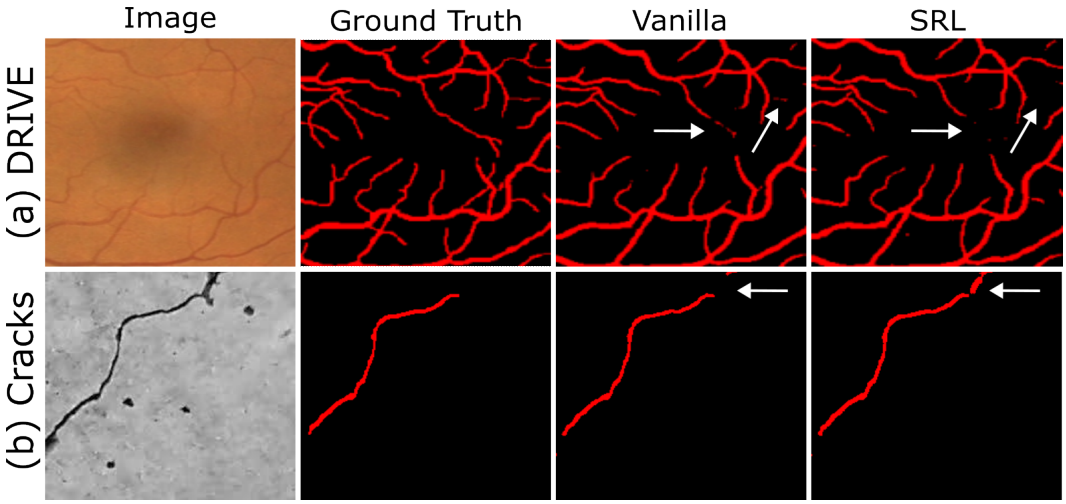


Figure 2: Results of proposed method over 2 of the datasets. nnUNet with conventional segmentation losses performs adequately in mapping the basic structure of the object of interest, oftenly outperforming the SRL method, as demonstrated in examples from DRIVE (a), Cracks (b) datasets.

Table 1: Test set metrics of nnUNet models trained on tubular datasets using different loss functions.

Method	**DSC** ↑	**clDice** ↑	**JSI** ↑	**FNR** ↓	**FPR** ↓
Cracks Dataset (Tomaszkiewicz & Owerko, 2023)					
Dice + CCE	**76.67**	**85.74**	**65.63**	20.38	**0.24**
Dice + CCE + SRL	75.57	85.43	64.66	**18.28**	0.32
DRIVE Dataset (Hassan et al., 2015)					
Dice + CCE	**84.02**	**87.85**	**72.48**	**19.22**	**1.55**
Dice + CCE + SRL	83.79	87.64	72.14	19.34	1.60

To investigate the gradient flowing through the network during back propagation, let us consider the gradient of each loss term with respect to the pixels in the predicted mask. The gradient of loss with respect to the ith pixel of predicted mask for the kth class, s_θ^{jk}, which is in turn dependent on parameter θ is given by:

$$\frac{\partial \mathscr{L}}{\partial s_\theta^{jk}} = \frac{\partial \mathscr{L}_{DCE}}{\partial s_\theta^{jk}} + \frac{\partial \mathscr{L}_{CE}}{\partial s_\theta^{jk}} + \frac{\partial \mathscr{L}_{SRL}}{\partial s_\theta^{jk}}$$

Given equation 1, further analyzing the gradient for the $\mathscr{L}_{SRL}$ term w.r.t. to any predicted pixel s_θ^{jk} for SRL is given by:

$$\frac{\partial \mathscr{L}_{SRL}}{\partial s_\theta^{jk}} = -\frac{1}{|K|}\frac{y^{jk}}{\sum_{i\in\Omega} y^{ik}} \tag{2}$$

From 2, it may be noted that the gradient for SRL with respect to the predicted mask is indeed independent of the predicted mask and depends only on the ground truth mask, which does not change while training the model. This gradient, therefore, remains constant throughout the training process, and hence keeps pushing the overall loss gradient in the same (constant) direction. $\mathscr{L}_{DCE}$ and $\mathscr{L}_{CE}$, on the other hand, keep altering the gradient direction based on model predictions. A deeper insight into the effect of SRL on the overall loss gradient is discussed in the appendix.

Therefore, we can conclude that the gradient of SRL just pushes the net gradient in an unnecessary constant direction over the epochs, in effect reducing the training efficiency.

REFERENCES

Waleed Abu-Ain, Siti Norul Huda Sheikh Abdullah, Bilal Bataineh, Tarik Abu-Ain, and Khairuddin Omar. Skeletonization algorithm for binary images. *Procedia Technology*, 11:704–709, 2013.

Olivier Bernard, Alain Lalande, Clement Zotti, Frederick Cervenansky, Xin Yang, Pheng-Ann Heng, Irem Cetin, Karim Lekadir, Oscar Camara, Miguel Angel Gonzalez Ballester, et al. Deep learning techniques for automatic mri cardiac multi-structures segmentation and diagnosis: is the problem solved? *IEEE transactions on medical imaging*, 37(11):2514–2525, 2018.

Mingfei Cheng, Kaili Zhao, Xuhong Guo, Yajing Xu, and Jun Guo. Joint topology-preserving and feature-refinement network for curvilinear structure segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 7147–7156, 2021.

Gehad Hassan, Aboul Ella Hassanien, Nashwa El-Bendary, and Ali Fahmy. Blood vessel segmentation approach for extracting the vasculature on retinal fundus images using particle swarm optimization. In *2015 11th international computer engineering conference (ICENCO)*, pp. 290–296. IEEE, 2015.

Fabian Isensee, Jens Petersen, Andre Klein, David Zimmerer, Paul F Jaeger, Simon Kohl, Jakob Wasserthal, Gregor Koehler, Tobias Norajitra, Sebastian Wirkert, et al. nnu-net: Self-adapting framework for u-net-based medical image segmentation. *arXiv preprint arXiv:1809.10486*, 2018.

Xun Jin and Jongweon Kim. A 3d skeletonization algorithm for 3d mesh models using a partial parallel 3d thinning algorithm and 3d skeleton correcting algorithm. *Applied Sciences*, 7(2):139, 2017.

Yannick Kirchhoff, Maximilian R Rokuss, Saikat Roy, Balint Kovacs, Constantin Ulrich, Tassilo Wald, Maximilian Zenk, Philipp Vollmuth, Jens Kleesiek, Fabian Isensee, et al. Skeleton recall loss for connectivity conserving and resource efficient segmentation of thin tubular structures. *arXiv preprint arXiv:2404.03010*, 2024.

Longin Jan Latecki, Quan-nan Li, Xiang Bai, and Wen-yu Liu. Skeletonization using ssm of the distance transform. In *2007 IEEE International Conference on Image Processing*, volume 5, pp. V–349. IEEE, 2007.

Martin J Menten, Johannes C Paetzold, Veronika A Zimmer, Suprosanna Shit, Ivan Ezhov, Robbie Holland, Monika Probst, Julia A Schnabel, and Daniel Rueckert. A skeletonization algorithm for gradient-based optimization, 2023.

Volodymyr Mnih. *Machine learning for aerial image labeling*. University of Toronto (Canada), 2013.

Panav Raina, Satyender Dharamdasani, Dheeraj Chinnam, Praveen Sharma, and Sukrit Gupta. Bombr: An annotated bone marrow biopsy dataset for segmentation of reticulin fibers. *bioRxiv*, pp. 2024–10, 2024.

Suprosanna Shit, Johannes C Paetzold, Anjany Sekuboyina, Ivan Ezhov, Alexander Unger, Andrey Zhylka, Josien PW Pluim, Ulrich Bauer, and Bjoern H Menze. cldice-a novel topology-preserving loss function for tubular structure segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 16560–16569, 2021.

Karolina Tomaszkiewicz and Tomasz Owerko. A pre-failure narrow concrete cracks dataset for engineering structures damage classification and segmentation. *Scientific Data*, 10(1):925, 2023.

Kaiyuan Yang, Fabio Musio, Yihui Ma, Norman Juchler, Johannes C Paetzold, Rami Al-Maskari, Luciano Höher, Hongwei Bran Li, Ibrahim Ethem Hamamci, Anjany Sekuboyina, et al. Benchmarking the cow with the topcow challenge: Topology-aware anatomical segmentation of the circle of willis for cta and mra. *ArXiv*, pp. arXiv–2312, 2024.