The Centerline-Cross Entropy Loss for Vessel-Like Structure Segmentation: Better Topology Consistency Without Sacrificing Accuracy

Cesar Acebes[1,2(✉)], Abdel Hakim Moustafa[2,3], Oscar Camara[1], and Adrian Galdran[1,4]

[1] Dept. of Information and Communication Technologies, Universitat Pompeu Fabra, Barcelona, Spain. cesar.acebes@upf.edu
[2] Dimension Lab, Hospital de la Santa Creu i Sant Pau, Barcelona, Spain.
[3] Cardiology Dept., Hospital de la Santa Creu i Sant Pau, Barcelona, Spain.
[4] Computer Vision Center, Universitat Autònoma de Barcelona, Spain.

Abstract. Achieving accurate vessel segmentation in medical images is crucial for various clinical applications, but current methods often struggle to balance topological consistency (preserving vessel network structure) with segmentation accuracy (overlap with ground-truth). Although various strategies have been proposed to address this challenge, they typically necessitate significant modifications to network architecture, more annotations, or entail prohibitive computational costs, providing only partial topological improvements. The clDice loss was recently proposed as an elegant and efficient alternative to preserve topology in tubular structure segmentation. However, segmentation accuracy is penalized and it lacks robustness to noisy annotations, mirroring the limitations of the conventional Dice loss. This work introduces the centerline-Cross Entropy (clCE) loss function, a novel approach which capitalizes on the robustness of Cross-Entropy loss and the topological focus of centerline-Dice loss, promoting optimal vessel overlap while maintaining faithful network structure. Extensive evaluations on diverse publicly available datasets (2D/3D, retinal/coronary) demonstrate clCE's effectiveness. Compared to existing losses, clCE achieves superior overlap with ground truth while simultaneously improving vascular connectivity. This paves the way for more accurate and clinically relevant vessel segmentation, particularly in complex 3D scenarios. We share an implementation of the clCE loss function in github.com/cesaracebes/centerline_CE.

Keywords: vessel segmentation · topology-preserving losses

1 Introduction

How do we estimate the performance of an image segmentation model? The conventional approach is to compute the overlap between target and predicted volumes, but the community is becoming increasingly aware that measuring performance of segmentation models is not as straightforward as it seems [19].

For example, it is well-known that when the foreground structures are small, overlap measures like the Dice Similarity Coefficient (DSC) become unstable and extremely sensitive to small changes in predictions [26]. Also, if targets of interest represent objects of varying sizes [17], if it is relevant to differentiate between under and over-segmentation errors [22], or in instance segmentation scenarios [13], overlap measures can be problematic.

Alternative segmentation measures include generalized overlap [3], distance metrics like Hausdorff distance [16], (scalar) volume similarity [22], or topological correctness [27]. Unfortunately, there is no silver bullet, as any metric will have its own limitations [26]. However, reflecting on measuring performance is not only relevant for assessing model quality and comparing between competing solutions, but also because the kind of performance metric we consider typically dictates the class of learning approach to adopt, and in particular the loss function to be optimized [16,27]. In addition, considering downstream applications when designing medical image segmentation models is crucial, since maximizing simple overlap might not be enough for clinical purposes.

In this paper, we address the problem of how to improve topological consistency of segmentation models. Applications for topology-preserving segmentation are manifold, including coronary artery segmentation [32], delineation of the Circle of Willis [31], rib segmentation [15] or retinal vasculature analysis [1], to name a few. All these challenging segmentation problems have attracted much research attention in recent years, with researchers proposing learning strategies to drive model optimization towards topologically consistent solutions. Some techniques rely on complementary annotations like adjacent anatomy [30,33], or vascular biomarkers [34]. Other approaches include adversarial learning [29], model ensembling [24], or model cascading [18]. While all these techniques can result in certain performance improvements, they require substantial modifications to network architectures or the availability of additional manual annotations. A more reasonable strategy is to build loss functions that encourage topological consistency in the learning process of segmentation models itself. For instance, loss functions inspired in topological properties like persistent homology can provide such benefits, but they are computationally intensive and challenging to implement, specially in 3D [4,11,28]. A popular alternative are loss functions that reward a model when it produces topologically consistent segmentations across its 2D projections [8,23], but it is a very application-dependent strategy.

A computationally efficient topology-inspired loss function that stands out is the centerline-Dice loss (clDice, [27]), with a straightforward formulation and implementation that has favoured its adoption. However, being based on the Dice segmentation loss, clDice inherits some of its negative properties, *e.g.* lack of robustness with respect to noisy annotations. In addition, by design clDice boosts topological consistency at the expense of segmentation accuracy [27].

Unlike the Dice segmentation loss, the Cross-Entropy (CE) loss is usually considered a more robust choice [6]; best practices recommend the combination of both losses for optimizing model performance. In this paper, we introduce the centerline-Cross Entropy (clCE) loss function, which is designed to generate seg-

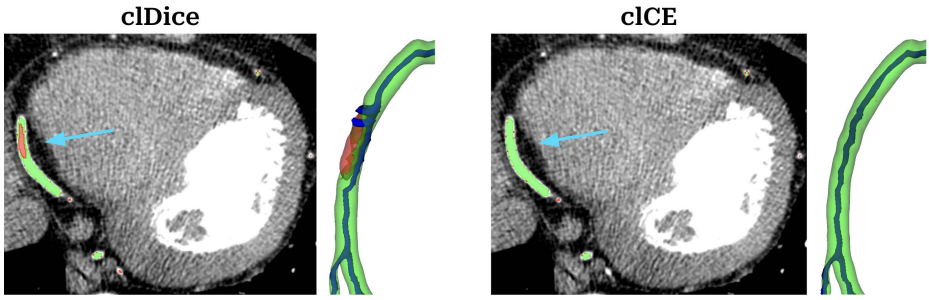

Fig. 1: Coronary artery segmentations using clDice and clCE losses in the ASOCA database. For each loss: Left: 2D image slice of the heart, highlighting the artery with an arrow. Right: 3D visualization of the resulting segmentations, showcasing the extracted centerline (dark blue). Green: TPs. Red: FPs and FNs.

mentations maximizing structure overlap, while simultaneously enhancing topological consistency, as illustrated in Fig. 1. The proposed approach was tested on a large variety of Open Access vessel segmentation datasets, including 2D and 3D problems in retinal and coronary artery applications.

2 Methodology

2.1 Definitions and Notation

For simplicity, consider a binary segmentation model, $\mathcal{U}_\theta$, trained to delineate a foreground structure; each image $\mathbf{x}$ has an associated ground-truth $\mathbf{T}$ of the same spatial size as $\mathbf{x}$ but taking values in $\{0, 1\}$. The standard goal for $\mathcal{U}_\theta$ is to maximize overlap between (binarized) predictions $\mathbf{P}$ and ground-truth $\mathbf{T}$. Overlap is typically measured as the ratio between $|\mathbf{T} \cap \mathbf{P}|$ and $|\mathbf{T} \cup \mathbf{P}|$: if both coincide, the segmentation is perfect. This observation gives raise to the Intersection over Union performance metric, and the closely related Dice Similarity Coefficient:

$$\mathrm{DSC}(\mathbf{T}, \mathbf{P}) = \frac{2|\mathbf{T} \cap \mathbf{P}|}{|\mathbf{T}| + |\mathbf{P}|} = \frac{2\mathrm{TP}}{2\mathrm{TP} + \mathrm{FP} + \mathrm{FN}} = \frac{\mathrm{Precision} \times \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}}, \quad (1)$$

which is the harmonic mean between Precision and Recall. Here, TP stands for True Positives, and FP/FN are False Positives/Negatives. From Equation (1) we see that the DSC does not account for True Negatives, which is considered as a beneficial property in medical image segmentation, where many pixels are easily identified as background and can bias the score towards artificially high values.

The next step involves formulating a loss function that can take non-binary predictions, $\mathcal{U}_\theta(\mathbf{x}) = \widehat{\mathbf{P}} \in [0, 1]$, compare them to $\mathbf{T}$ and return a positive number $\mathcal{L}(\mathbf{T}, \widehat{\mathbf{P}})$ that becomes lower as the prediction improves. A direct generalization of Equation (1) leads to the Dice loss $\mathcal{L}_{Dice}$, with the Binary Cross-Entropy loss

$\mathcal{L}_{CE}$ being a popular alternative:

$$\mathcal{L}_{Dice}(\mathbf{T}, \widehat{\mathbf{P}}) = 1 - \frac{2\langle \mathbf{T}, \widehat{\mathbf{P}} \rangle}{\langle \mathbf{T}, \mathbf{T} \rangle + \langle \widehat{\mathbf{P}}, \widehat{\mathbf{P}} \rangle}, \tag{2}$$

$$\mathcal{L}_{CE}(\mathbf{T}, \widehat{\mathbf{P}}) = -\sum_i \mathbf{T}_i \log(\widehat{\mathbf{P}}_i) + (1 - \mathbf{T}_i) \log(1 - \widehat{\mathbf{P}}_i), \tag{3}$$

where i indexes image voxels. Both losses have complementary properties, with no consensus on which one is optimal for any given problem, as discussed next.

2.2 Advantages and drawbacks of standard segmentation losses

It is well-known that the DSC, and by extension the Dice loss, suffer from a bias towards small targets [17,19]. In consequence, if in the course of a batch-wise iterative optimization process, a particular batch contains a small foreground with very few voxels being wrongly predicted, the loss of this sample will dominate over the others. The same applies when performance is measured on a set of images and averaged over the individual DSCs, which may be problematic due to typical annotation ambiguities around foreground borders in medical imaging.

On the other hand, the CE loss is often preferred when robustness is required [20]. Thanks to its "regional" character (it considers all voxels) and its better calibration (it heavily penalizes extremely confident and wrong predictions), its minimization may result in more general solutions [6]. Also, a linear combination of both losses is sometimes preferred, as in the popular nnUNet framework [12].

Yet another troublesome situation arises when it is beneficial to evaluate a segmentation as better if it does not break foreground connectivity, *i.e.*, if it preserves the topological structure of the object to segment, even if a few of its voxels are missed. This is often the case for foreground objects of tubular shape, *e.g.* in vascular segmentation problems. In this case, neither Dice nor the CE loss reflect relevant properties of an ideal solution.

2.3 Topology-preserving metrics and losses

In a context where it might be more relevant to preserve vessel connectivity, even if the exact contour is not captured, topology-preserving measures need to be considered. The most common way to reward topology preservation is by carrying out a skeletonization process on both the target $\mathbf{T}$ and the binary prediction $\mathbf{P}$, obtaining the corresponding skeletons (or centerlines) $\mathbf{S_T}$ and $\mathbf{S_P}$, and then calculating some measure of similarity between them. However, it is not a straightforward process since two close-by skeletons might share little overlap and still be both valid segmentations.

The metric of choice for tubular-like segmentation problems is the centerline Dice coefficient (clDice, [27]), which is computed by counting all voxels from $\mathbf{S_P}$ inside the target volume $\mathbf{T}$, and then all voxels in $\mathbf{S_T}$ within the predicted volume $\mathbf{P}$. This gives raise to the definition of *Topology-Precision* and *Topology-Recall*:

$$\mathcal{T}_{\text{prec}}(\mathbf{T}, \mathbf{P}) = \frac{|\mathbf{T} \cap \mathbf{S_P}|}{|\mathbf{S_P}|}, \qquad \mathcal{T}_{\text{recall}}(\mathbf{T}, \mathbf{P}) = \frac{|\mathbf{P} \cap \mathbf{S_T}|}{|\mathbf{S_T}|}. \tag{4}$$

The clDice metric is then defined by analogy to the DSC as the harmonic mean between $\mathcal{T}_{\text{prec}}$ and $\mathcal{T}_{\text{recall}}$. In order to turn the clDice metric into a loss function, we need a way of extracting skeletons from non-binary predictions $\widehat{\mathbf{P}}$. This can be achieved with soft (gray-scale) morphological erosions and dilations [27]. As these operations can be written in terms of max-pooling operations, an iterative skeletonization algorithm can be easily implemented as a neural network layer.

2.4 The centerline-Cross Entropy loss

Our hypothesis is that the clDice loss suffer from similar shortcomings as the DSC: a lack of robustness to noise and variability in annotations. Since iterative skeletonization methods are noisy processes [21], a non-robust relying on imperfect skeletons loss may lead to finding suboptimal models. In order to overcome this limitation, we first need to consider a probabilistic interpretation of the topology precision and recall metrics in Equation (4), where $\mathbb{P}$ denotes probability:

$$\mathcal{T}_{\text{prec}}(\mathbf{T}, \mathbf{P}) = \mathbb{P}(\mathbf{T}|\mathbf{S_P}), \qquad \mathcal{T}_{\text{recall}}(\mathbf{T}, \mathbf{P}) = \mathbb{P}(\mathbf{P}|\mathbf{S_T}).$$

It can be seen that in order to optimize both quantities, we need to maximize the performance of the model conditioned on voxels belonging to both the skeletons of the target and the prediction. However, instead of mimicking the Dice loss mechanism, we propose to emulate the Cross-Entropy philosophy of attaching a negative logarithmic loss to wrong pixels. When conditioning on the skeleton of the target, we only need to average the voxel-wise CE loss over the target's skeleton $\mathbf{S_T}$:

$$\text{CE-}\mathcal{T}_{\text{prec}}(\mathbf{T}, \widehat{\mathbf{P}}) = -\frac{1}{\|\mathbf{S_T}\|_1} \mathcal{L}_{\text{CE}}(\mathbf{T}, \widehat{\mathbf{P}}) \odot \mathbf{S_T} = -\frac{1}{\|\mathbf{S_T}\|_1} \sum_{i \,|\, \mathbf{x}_i \in \mathbf{S_T}} \mathbf{T}_i \log(\widehat{\mathbf{P}}_i). \tag{5}$$

The part of our loss function that operates over voxels on the prediction's skeleton is less direct, since the skeleton needs to be computed from a non-binary image. We use the algorithm in [27] for computing a soft-skeleton $\mathbf{S}_{\widehat{\mathbf{P}}}$, and then we obtain a weighted average of the voxel-wise CE loss over that soft skeleton:

$$\begin{aligned} \text{CE-}\mathcal{T}_{\text{recall}}(\mathbf{T}, \widehat{\mathbf{P}}) &= -\frac{1}{\|\mathbf{S}_{\widehat{\mathbf{P}}}\|_1} \mathcal{L}_{\text{CE}}(\mathbf{T}, \widehat{\mathbf{P}}) \odot \mathbf{S}_{\widehat{\mathbf{P}}} \\ &= -\frac{\sum_{i \,|\, \mathbf{x}_i \in \mathbf{S}_{\widehat{\mathbf{P}}}} (\mathbf{T}_i \log(\widehat{\mathbf{P}}_i) + (1 - \mathbf{T}_i) \log(1 - \widehat{\mathbf{P}}_i)) \cdot \widehat{\mathbf{P}}_i}{\sum_{i \,|\, \mathbf{x}_i \in \mathbf{S}_{\widehat{\mathbf{P}}}} \widehat{\mathbf{P}}_i}. \end{aligned} \tag{6}$$

```
import torch, torch.nn.functional as F

def clCE(y_pred, y_true):
    l_ce = F.cross_entropy(y_pred, y_true, reduction='none')
    y_pred = y_pred.softmax(dim=1)
    y_true_sk = soft_skel(y_true)
    y_pred_sk = soft_skel(y_pred)
    ce_tprec = torch.mul(l_ce, y_true_sk[:, 1]).mean()
    ce_trecall = torch.mul(l_ce, y_pred_sk[:, 1]).mean()
    return ce_tprec+ce_trecall
```

Code 1: Implementation of the centerline-Cross Entropy (clCE) loss; `soft_skel` computes a soft skeleton.

In these expressions, $\odot$ represents the voxel-wise product. Finally, we compute the sum of both terms, arriving at the definition of the centerline-CE loss (clCE):

$$\mathcal{L}_{\text{clCE}}(\mathbf{T}, \widehat{\mathbf{P}}) = \text{CE-}\mathcal{T}_{\text{prec}}(\mathbf{T}, \widehat{\mathbf{P}}) + \text{CE-}\mathcal{T}_{\text{recall}}(\mathbf{T}, \widehat{\mathbf{P}}). \tag{7}$$

Since both operations in Equations (5) and (6) can be expressed as a weighted average, the resulting implementation consists of a few lines of code, see 1.

3 Experimental analysis

3.1 Datasets, models, and performance evaluation

In order to validate the proposed clCE loss, we carry out experiments on a variety of datasets, all involving vasculature segmentation: 1) The HRF dataset (link) contains 45 eye fundus images with a resolution of 3504×2336 and 45° field of view [2]. 2) The DR-HAGIS database (link) has 40 retinal images with 45° FOV from a variety of devices and resolutions, ranging from 2816×1880 to 4752×3168 [10]. 3) The TREND database (link) has 82 retinal images acquired with a hand-held portable camera, at 45º FOV and 2560×1920 pixel resolution [25]. 4) The FIVES dataset (link) has 800 retinal images with a 50° FOV captured at 2048×2048 resolution [14]. FIVES is divided into: FIVES-N (healthy eyes), FIVES-G (glaucomatous eyes), FIVES-A (age-related macular degeneration) and FIVES-D (diabetic retinopathy), each with 200 samples. **5)** The ASOCA database (link) has 40 CCTA scans with 0.3-0.4 mm in-plane and 0.625 out-of-plane voxel size [7].

All considered datasets include images from pathological cases and healthy controls. We deliberately avoid experimenting on legacy low-resolution retinal datasets like DRIVE, as this is not realistic in current clinical practice. On the retinal datasets, we employ a 60/20/20 train/validation/test split. For the FIVES dataset, we train only on non-pathological images, and then test separately on

Table 1: Results on the HRF, DR-HAGIS, TREND, and ASOCA datasets. Best performance underlined. Green is an **increase** from baseline performance, red a **decrease**.

	$\mathcal{L}_{Dice}$		$\mathcal{L}_{Dice} + 0.5 \cdot \mathcal{L}_{clDice}$		$\mathcal{L}_{Dice} + \mathcal{L}_{clCE}$	
	DSC	**cl-DSC**	**DSC**	**cl-DSC**	**DSC**	**cl-DSC**
HRF	**78.93**	**77.46**	**75.60** (-3.33)	**<u>83.83</u>** (+6.37)	**<u>79.34</u>** (+0.41)	**78.94** (+1.48)
DR-HAGIS	**69.22**	**68.73**	**66.62** (-2.60)	**<u>74.20</u>** (+5.47)	**<u>70.65</u>** (+1.43)	**70.06** (+1.33)
TREND	**63.92**	**65.32**	**61.04** (-2.88)	**<u>71.35</u>** (+6.03)	**<u>64.54</u>** (+0.62)	**67.04** (+1.72)
ASOCA	**84.59**	**84.81**	**83.42** (-1.17)	**84.76** (-0.05)	**<u>84.80</u>** (+0.21)	**<u>84.95</u>** (+0.14)

each sub-dataset. In the ASOCA database, since the test set is withheld by challenge organizers, we split the data and run a 5-fold cross-validation analysis.

For training in 2D we use the small U-Net proposed in [5], as it was shown capable of reaching state-of-the-art results while enabling fast experimentation. During training, we extract 512×512 patches from the images at native resolution, and use a sliding window approach in test time. For training in 3D, the nnUNet framework from [12] is used in its full-resolution 3D configuration, only modifying the loss function. Exact training details are available at `github.com/cesaracebes/centerline_CE`.

For comparison purposes, we optimize each model minimizing first the $\mathcal{L}_{Dice}$ loss, and then using $\mathcal{L}_{Dice} + \mathcal{L}_{clDice}$, which is the setup proposed in the original work. However, this loss function showed a very unstable behaviour, which was fixed by halving the contribution of $\mathcal{L}_{clDice}$. We then train with $\mathcal{L}_{Dice} + \mathcal{L}_{clCE}$, and compute the DSC for analyzing segmentation performance, with the clDice metric assessing topological consistency[1]. It is expected that the $\mathcal{L}_{clDice}$ will lead to improvements in clDice metric, by sacrificing segmentation accuracy in terms of DSC, as observed in [27]. Let us stress that we do not expect the clCE loss to outperform the clDice loss in terms of the clDice metric, but rather to improve topological consistency without sacrificing segmentation performance.

3.2 Numerical results

First, results on all retinal datasets, unless FIVES, are provided on Table 1, together with the ASOCA experiment. As expected, in most cases training with the Dice loss combined with either $\mathcal{L}_{clDice}$ or $\mathcal{L}_{clCE}$ leads to noticeable improvements in topological consistency. It can also be observed that using $\mathcal{L}_{clDice}$ results in larger clDice metric increases, albeit at the cost of segmentation accuracy, as evidenced by the decrease in DSC across the board for this loss. In contrast,

[1] Supplementary materials contain results with means and deviations.

Table 2: Results on FIVES. Models trained on FIVES-N (non-pathological). A, G, D indicates age-related macular degeneration, glaucomatous eyes and diabetic retinopathy (pathological), respectively. Best performance underlined. Green means an **increase** from baseline performance, and red a **decrease**.

	$\mathcal{L}_{Dice}$		$\mathcal{L}_{Dice} + 0.5 \cdot \mathcal{L}_{clDice}$		$\mathcal{L}_{Dice} + \mathcal{L}_{clCE}$	
	DSC	**cl-DSC**	**DSC**	**cl-DSC**	**DSC**	**cl-DSC**
FIVES-N	**83.06**	**80.29**	**77.14** (-5.92)	**<u>83.37</u>** (+3.08)	**<u>83.40</u>** (+0.34)	**81.28** (+0.99)
FIVES-A	**83.81**	**79.50**	**80.80** (-3.01)	**<u>84.48</u>** (+4.98)	**<u>84.10</u>** (+0.29)	**80.64** (+1.14)
FIVES-G	**71.75**	**66.88**	**66.03** (-5.72)	**67.81** (+0.93)	**<u>71.83</u>** (+0.08)	**<u>67.91</u>** (+1.03)
FIVES-D	**78.62**	**73.35**	**74.59** (-4.03)	**<u>79.03</u>** (+5.68)	**<u>79.04</u>** (+0.42)	**74.65** (+1.3)

training with $\mathcal{L}_{clCE}$ leads to increases in both DSC and clDice measures everywhere. It is also interesting to notice that, for the most complex scenario, *i.e.* the 3D segmentation problem on the ASOCA dataset, the $\mathcal{L}_{clDice}$ not only attains lower segmentation performance, but it is unable to improve the topological consistency of the baseline $\mathcal{L}_{Dice}$. Also in this case, the $\mathcal{L}_{clCE}$ can enhance results in both aspects of the problem. Note that we also experimented with replacing $\mathcal{L}_{Dice}$ with the $\mathcal{L}_{CE}$ as a baseline, and in combination with $\mathcal{L}_{clDice}/\mathcal{L}_{clCE}$, observing exactly the same trends. These results are included in the supplement.

The experiment in Table 2 was designed to verify the hypothesis that the $\mathcal{L}_{clCE}$ inherits the robustness of $\mathcal{L}_{CE}$. Models trained on healthy subjects were tested on both healthy and pathological images, which results in sub-group performance declines. However, also in this case we observe a similar trend as above: the $\mathcal{L}_{clCE}$ is capable of increasing in all cases both the DSC and the clDice metric, while the $\mathcal{L}_{clDice}$ also improves topological consistency, but fails to reach the segmentation performance of training just with $\mathcal{L}_{Dice}$.

4 Conclusions

When segmenting vessel-like structures, it is often a requirement for downstream tasks that the vascular topology is preserved as faithfully as possible. Existing topology-preserving losses encourage that behaviour, but they achieve it at the expense of subpar overall segmentation performance. A simple solution would be to train two different models, one specialized on topological properties and then another one on generic segmentation; however, this would be a costly approach, unfeasible in large 3D segmentation problems. A more reasonable strategy is to seek loss functions that can improve segmentation in terms of ground-truth overlap, while also preserving topological properties of the target.

In this paper we have introduced the centerline-Cross Entropy (clCE), a new loss function that promotes topology-preserving solutions without sacrificing segmentation accuracy. Our experiments consistently indicate that the clCE loss achieves better overlap than existing losses, and also manages to improve vascular connectivity, thereby offering a competitive solution that can be advantageous when segmenting complex datasets. Future work will be devoted to compute and learn better skeleton representation, such as with differentiable methods [21], and graph-neural networks [9].

Disclosure of interests The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Araújo, R.J., et al.: A Deep Learning Design for Improving Topology Coherence in Blood Vessel Segmentation. In: MICCAI 2019. pp. 93–101 (2019). https://doi.org/10.1007/978-3-030-32239-7$_1$1 2
2. Budai, A., et al.: Robust Vessel Segmentation in Fundus Images. International Journal of Biomedical Imaging (2013). https://doi.org/10.1155/2013/154860 6
3. Crum, W., et al.: Generalized Overlap Measures for Evaluation and Validation in Medical Image Analysis. IEEE Transactions on Medical Imaging **25**(11), 1451–1461 (Nov 2006). https://doi.org/10.1109/TMI.2006.880587 2
4. Fu, G., et al.: Introducing soft topology constraints in deep learning-based segmentation using projected pooling loss. In: Medical Imaging 2023: Image Processing. vol. 12464 (Apr 2023). https://doi.org/10.1117/12.2651576 2
5. Galdran, A., et al.: State-of-the-art retinal vessel segmentation with minimalistic models. Scientific Reports **12**(1) (Apr 2022). https://doi.org/10.1038/s41598-022-09675-y 7
6. Galdran, A., et al.: On the Optimal Combination of Cross-Entropy and Soft Dice Losses for Lesion Segmentation with Out-of-Distribution Robustness. In: Diabetic Foot Ulcers Grand Challenge, pp. 40–51 (2023) 2, 4
7. Gharleghi, R., et al.: Automated segmentation of normal and diseased coronary arteries – The ASOCA challenge. Computerized Medical Imaging and Graphics **97** (Apr 2022). https://doi.org/10.1016/j.compmedimag.2022.102049 6
8. Guo, Z., et al.: 3D Vascular Segmentation Supervised by 2D Annotation of Maximum Intensity Projection. IEEE Transactions on Medical Imaging (2024). https://doi.org/10.1109/TMI.2024.3362847 2
9. He, H., et al.: Automated Coronary Vessels Segmentation in X-ray Angiography Using Graph Attention Network, p. 209–219. Springer Nature Switzerland (2024). https://doi.org/10.1007/978-3-031-52448-6$_2$0 9
10. Holm, S., et al.: DR HAGIS-a fundus image database for the automatic extraction of retinal surface vessels from diabetic patients. Journal of Medical Imaging **4**(1) (Jan 2017). https://doi.org/10.1117/1.JMI.4.1.014503 6
11. Hu, X., et al.: Topology-Preserving Deep Image Segmentation. In: Neural Information Processing Systems (2019) 2
12. Isensee, F., et al.: nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nature Methods **18**(2), 203–211 (Feb 2021). https://doi.org/10.1038/s41592-020-01008-z 4, 7

13. Jena, R., et al.: Beyond mAP: Towards Better Evaluation of Instance Segmentation. In: Computer Vision and Pattern Recognition (Jun 2023) 2
14. Jin, K., et al.: FIVES: A Fundus Image Dataset for Artificial Intelligence based Vessel Segmentation. Scientific Data **9**(1), 475 (Aug 2022). https://doi.org/10.1038/s41597-022-01564-3 6
15. Jin, L., et al.: RibSeg v2: A Large-Scale Benchmark for Rib Labeling and Anatomical Centerline Extraction. IEEE Transactions on Medical Imaging **43**(1) (Jan 2024). https://doi.org/10.1109/TMI.2023.3313627 2
16. Karimi, D., et al.: Reducing the Hausdorff Distance in Medical Image Segmentation With Convolutional Neural Networks. IEEE Transactions on Medical Imaging **39**(2), 499–513 (Feb 2020). https://doi.org/10.1109/TMI.2019.2930068 2
17. Liu, B., et al.: Do we really need dice? The hidden region-size biases of segmentation losses. Medical Image Analysis **91** (Jan 2024). https://doi.org/10.1016/j.media.2023.103015 2, 4
18. Luo, S., et al.: Two-Stage Topological Refinement Network for Retinal Artery/Vein Classification. In: ISBI. pp. 1–4 (Mar 2022). https://doi.org/10.1109/ISBI52829.2022.9761669, iSSN: 1945-8452 2
19. Maier-Hein, L., et al.: Metrics reloaded: Recommendations for image analysis validation. Nature Methods **21**(2) (Feb 2024). https://doi.org/10.1038/s41592-023-02151-z 1, 4
20. Mehrtash, A., et al.: Confidence Calibration and Predictive Uncertainty Estimation for Deep Medical Image Segmentation. IEEE Transactions on Medical Imaging **39**(12) (Dec 2020). https://doi.org/10.1109/TMI.2020.3006437 4
21. Menten, M.J., et al., Rueckert, D.: A skeletonization algorithm for gradient-based optimization. In: Computer Vision and Pattern Recognition (Oct 2023). https://doi.org/10.1109/ICCV51070.2023.01956, iSSN: 2380-7504 5, 9
22. Nai, Y.H., et al.: Comparison of metrics for the evaluation of medical segmentations using prostate MRI dataset. Computers in Biology and Medicine **134** (Jul 2021). https://doi.org/10.1016/j.compbiomed.2021.104497 2
23. Oner, D., et al.: Enforcing Connectivity of 3D Linear Structures Using Their 2D Projections, p. 591–601. Springer Nature Switzerland (2022). https://doi.org/10.1007/978-3-031-16443-9$_5$7 2
24. Park, J., et al.: Selective ensemble methods for deep learning segmentation of major vessels in invasive coronary angiography. Medical Physics **50**(12) (2023). https://doi.org/10.1002/mp.16554 2
25. Popovic, N., et al., S.: TREND database: Retinal images of healthy young subjects visualized by a portable digital non-mydriatic fundus camera. PLoS ONE **16**(7) (Jul 2021). https://doi.org/10.1371/journal.pone.0254918 6
26. Reinke, A., et al.: Understanding metric-related pitfalls in image analysis validation. Nature Methods **21**(2) (Feb 2024). https://doi.org/10.1038/s41592-023-02150-0 2
27. Shit, S., et al.: clDice - a Novel Topology-Preserving Loss Function for Tubular Structure Segmentation. In: Computer Vision and Pattern Recognition (Jun 2021). https://doi.org/10.1109/CVPR46437.2021.01629, iSSN: 2575-7075 2, 4, 5, 7
28. Stucki, N., et al.: Topologically faithful image segmentation via induced matching of persistence barcodes. In: ICML (2023) 2
29. Tang, Z., et al.: Adversarial Transformer for Repairing Human Airway Segmentation. IEEE Journal of Biomedical and Health Informatics **27**(10) (Oct 2023). https://doi.org/10.1109/JBHI.2023.3290136 2

30. Wang, W., et al.: AVDNet: Joint coronary artery and vein segmentation with topological consistency. Medical Image Analysis **91**, 102999 (Jan 2024). https://doi.org/10.1016/j.media.2023.102999 2
31. Yang, K., et al.: TopCoW: Benchmarking Topology-Aware Anatomical Segmentation of the Circle of Willis (CoW) for CTA and MRA (Jan 2024). https://doi.org/10.48550/arXiv.2312.17670 2
32. Zhang, X., et al.: Progressive Deep Segmentation of Coronary Artery via Hierarchical Topology Learning. In: MICCAI 2022 (2022). https://doi.org/10.1007/978-3-031-16443-$9_3$8 2
33. Zhang, X., et al.: An Anatomy- and Topology-Preserving Framework for Coronary Artery Segmentation. IEEE Transactions on Medical Imaging **43**(2) (Feb 2024). https://doi.org/10.1109/TMI.2023.3319720 2
34. Zhou, Y., et al.: CF-Loss: Clinically-relevant feature optimised loss function for retinal multi-class vessel segmentation and vascular feature measurement. Medical Image Analysis **93** (Apr 2024). https://doi.org/10.1016/j.media.2024.103098 2