

Smooth cLDice: A Reliable Metric for Vascular Segmentation Evaluation

oscar.morand
(supervisor: Elodie Puybareau)

January 2025

Vascular segmentation is a key task in medical image analysis, essential for diagnosis, monitoring, and the planning of interventions and treatments. However, it poses significant challenges, such as variability in annotations among experts, topology, and the lack of relevant metrics that adequately capture the topological aspects of these structures. cLDice was introduced as a metric to evaluate models' ability to preserve the topology of vascular structures and was used as a loss function to guide these models. However, this metric has limitations, particularly when dealing with very thin vessels and imprecise annotations, often caused by partial volume effects. To address these issues, we propose a variant of cLDice, the Smooth cLDice, which incorporates an uncertainty zone. This new metric allows for weighting errors at vessel boundaries while penalizing more heavily those that compromise the predicted vascular structure. We also introduce a differentiable version of this metric to enable its use as a loss function during model training.

La segmentation vasculaire est une tâche clé en analyse d'images médicales, essentielle pour le diagnostic, le suivi et la planification des interventions et des traitements. Cependant, elle présente des défis importants tels que la variabilité des annotations entre experts, la topologie, et le manque de métriques pertinentes, qui ne capturent pas toujours les aspects topologiques de ces structures. Le cLDice a été introduit comme métrique pour évaluer la capacité des modèles à préserver la topologie des structures vasculaires et utilisé comme fonction de coût pour guider ces modèles. Toutefois, cette métrique présente des limitations, notamment face aux vaisseaux très fins et aux annotations imprécises souvent dues à l'effet de volume partiel. Pour y remédier, nous proposons une variante du cLDice, le Smooth cLDice, intégrant une zone d'incertitude. Cette nouvelle métrique permet de pondérer les erreurs aux bordures des vaisseaux tout en pénalisant davantage celles qui compromettent la structure vasculaire prédite. Nous présentons aussi une version différentiable de cette métrique pour pouvoir l'utiliser comme fonction de coût lors d'entraînement de modèles.

Keywords

Vascular segmentation, Metric, Loss function, Deep Learning



Laboratoire de Recherche de l'EPITA
14-16, rue Voltaire
94270 Le Kremlin-Bicêtre CEDEX
France

Copying this document

Copyright © 2023 LRE.

Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.2 or any later version published by the Free Software Foundation; with the Invariant Sections being just “Copying this document”, no Front-Cover Texts, and no Back-Cover Texts.

A copy of the license is provided in the file COPYING.DOC.

Contents

1	Introduction	5
2	Problematics	7
2.1	The Partial Volume Effect	7
2.2	Variability in Annotations	8
2.3	Limitations of Current Metrics	9
3	Smooth cIDice, the metric	10
3.1	cIDice	10
3.2	Our addition to cIDice: Smooth cIDice	11
3.3	Verification of metric's behavior	12
3.4	Conclusion on the Smooth cIDice metric and future directions	14
4	A gradient-based optimization compatible version of Smooth cIDice	15
4.1	Conditions to adapt the metric as a loss function	15
4.2	The skeletonization algorithm	15
4.2.1	The skeletonization algorithm based on iterative pooling functions	16
4.3	Smoothing function, distance map algorithm	17
4.3.1	First iteration, simple algorithm	18
4.3.2	Second iteration, complex algorithm	18
4.4	Limits of the loss function, practical application	19
5	Experience: loss functions comparison	20
5.1	Experience protocol	20
5.1.1	Datasets Selection	20
5.1.2	Metric Selection	21
5.1.3	Cost Functions Selection	21
5.1.4	Model	22
5.1.5	Training	22
5.2	Results analysis	23
5.2.1	Global metrics results	23
5.2.2	Results by database	23
5.2.3	Ranking of loss functions	28
6	Future work	29
6.1	Supervision of the project	30
6.2	Metric coherence verification	30

CONTENTS	4
7 Conclusion	32
8 Bibliography	33

Chapter 1

Introduction

To fully understand the subject of this project and the challenges it aims to address, it is important to first clarify each key term in the title.

Vascular segmentation ([1.1](#)) is a crucial task in medical image analysis, essential for diagnosis, monitoring, and treatment planning. Specifically, this is a binary segmentation task, where the objective is to classify the pixels or voxels in an image into two categories: blood vessels (1) and the background (0), which includes all other tissues. Accurate vascular segmentation is critical to ensure precise medical interpretations, such as detecting abnormalities or planning medical interventions.

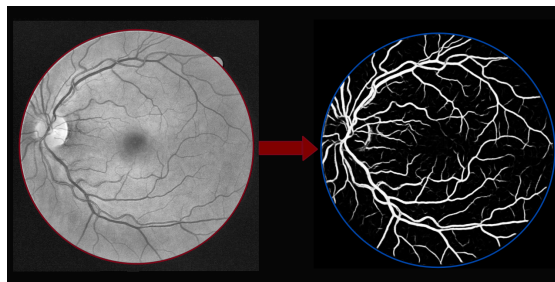


Figure 1.1: Example of vascular segmentation [Givisiez \(2025\)](#)

A metric is a measure used to evaluate the performance of a machine learning model on a specific task. Here, we emphasize the term “reliable” because no metric is universally superior, instead, metrics are designed to suit specific tasks. This work focuses on specialized metrics that assess a model’s ability to segment blood vessels while preserving their topology, which refers to the structure and connectivity of the vessels. This means that the model must segment blood vessels while maintaining their continuity, branching, and overall shape. For example, in a vascular network, it is critical that segments remain uninterrupted and properly connected. Errors in these aspects could lead to incorrect interpretations of vascular structures, which is particularly problematic in medical applications where decisions based on these analyses can have significant consequences.

While prioritizing the topology of vascular networks introduces a certain bias, it is important

to acknowledge that some medical applications may place higher importance on other parameters. For instance, there are cases where the characteristics of vessel walls may require more attention. However, recent research publications in this field ([Stucki et al. \(2023\)](#), [Shi et al. \(2024\)](#), [Acebes et al. \(2024\)](#)) increasingly support the idea that topological fidelity should be prioritized, highlighting its significance in medical imaging.

Lastly, Smooth cIDice is the name we have chosen for the new metric introduced in this paper. This metric represents an innovative approach to evaluating and guiding vascular segmentation models by addressing the limitations of existing methods and incorporating considerations for uncertainty and topological integrity.

This work is based on several public vascular datasets ([5.1.1](#)), including retinal fundus images, to ensure comprehensive and representative evaluation.

Chapter 2

Problematics

It is essential to grasp the challenges associated with the structures under study to better understand the solutions proposed by prior research and our own contributions.

2.1 The Partial Volume Effect

The partial volume effect (2.1) is a well-known issue present in all medical imaging modalities, stemming from the inherent nature of the data. This effect arises due to the need to represent continuous physical structures in a discrete format for imaging purposes.

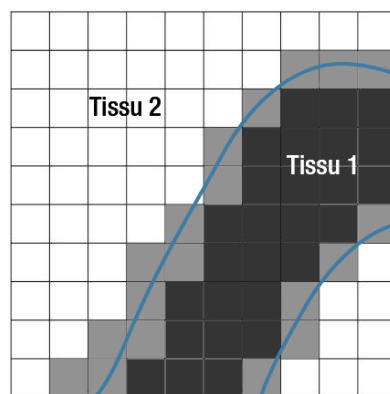


Figure 2.1: Example of the partial volume effect. The gray pixels correspond to partial volume artifacts and have an intermediate value between tissues 1 and 2.

While discretization generally functions adequately, problems emerge at the boundaries between tissues. A pixel located at the interface of two tissues with distinct imaging values, such as bone and muscle, will encode a value that is a combination of both tissues. Consequently, this pixel's intensity does not correspond accurately to either tissue.

This effect introduces a blurring at the tissue boundaries, making it challenging for experts or computational models to unambiguously classify such pixels. For large organs like the liver or

heart, this uncertainty is negligible since the ambiguous pixels represent only a small proportion of the overall structure.

However, in vascular segmentation, the impact of the partial volume effect is significant due to the unique characteristics of blood vessels, as they often occupy a minimal portion of the image or volume, particularly in high-resolution datasets. It is common to encounter vessels that are only a few pixels thick. In such cases, a substantial fraction of the vessel's pixels is affected by the partial volume effect, rendering these structures nearly undecidable in terms of their exact location within the image.

2.2 Variability in Annotations

This challenge is further compounded when observing manually created annotations by medical professionals. Consider, for example, two ground truths generated by different annotators for the same dataset (2.2), such as the STARE dataset Hoover and Goldbaum (2000). These ground truths should ideally be similar, with only minor differences at the boundaries due to the uncertainties introduced by the partial volume effect.

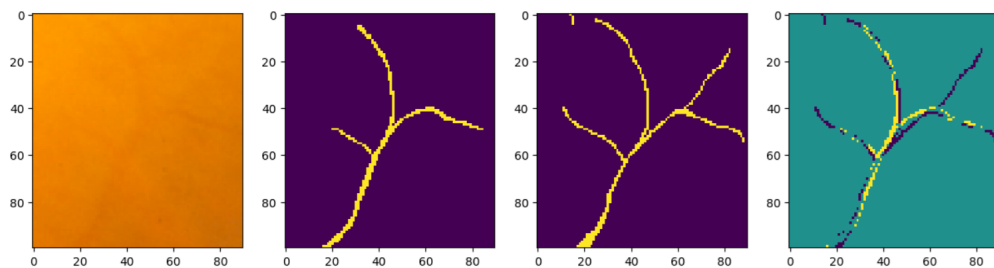


Figure 2.2: Image of the blood vessel (left), ground truth made by person number 1, ground truth number 2, map of the differences between the two ground truths (right).

Upon closer examination, however, the differences can be striking. By zooming into the same region and comparing the two ground truths, it becomes evident that they diverge significantly. Calculating the Dice score between these two ground truths yields a value of 0.45, which is notably low.

Although methods exist to merge ground truths to approximate an average representation, this highlights the considerable variability between annotations. Our objective is to design a metric that can account for and mitigate this variability. The metric must remain robust to annotator differences, particularly for small blood vessels. Ideally, it would assign a better score to the two ground truths mentioned, as they should both represent valid interpretations of the same structures.

It is important to emphasize that this does not reflect on the annotators' expertise. Segmentation of such intricate structures is inherently difficult and often impossible to perform with complete precision due to the challenges outlined above.

2.3 Limitations of Current Metrics

Finally, the combination of these challenges reveals significant shortcomings in existing metrics for vascular segmentation tasks. Most widely used metrics 2.4, such as accuracy or the F1 score, rely on pixel-wise comparisons between predicted and ground truth masks. These metrics, which calculate true positives, false negatives, and other values of the confusion matrix 2.3, based on the areas of the masks, are not well-suited to evaluating the quality of vascular segmentation, especially when dealing with small blood vessels.

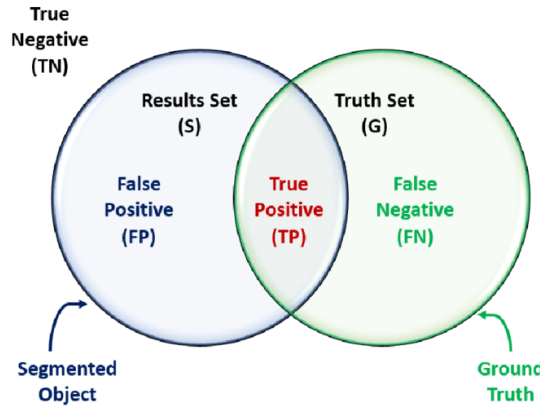


Figure 2.3: Diagram Mostapha (2014) illustrating the calculation of confusion matrix values for segmentation.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\text{Precision / Positive predictive value (PPV)} = \frac{TP}{TP + FP}$$

$$\text{Recall / Sensitivity} = \frac{TP}{TP + FN}$$

$$\text{Specificity} = \frac{TN}{TN + FP}$$

$$F1 = \frac{2 \times TP}{2 \times TP + FP + FN} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

$$\text{False Positive Rate / Fallout} = 1 - \text{Specificity} = \frac{FP}{TN + FP}$$

$$\text{Negative Predictive Value (NPV)} = \frac{TN}{TN + FN}$$

$$\text{MCC} = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

Figure 2.4: Example of classic metrics based on the calculation of confusion matrix values.

To address this, it is essential to move beyond traditional pixel-wise metrics and develop a more robust evaluation framework. The new metric must effectively handle the specific issues of partial volume effects, annotation variability, and the unique topology of vascular structures.

Chapter 3

Smooth cIDice, the metric

Our work is based on these problematics and builds upon a paper introducing cIDice [Shit et al. \(2021\)](#), a novel loss function designed to preserve topology in the segmentation of tubular structures.

3.1 cIDice

cIDice metric was developed to evaluate a model's ability to maintain the connectivity of vascular structures, making it particularly relevant to our research question.

Unlike traditional metrics such as the Dice score, which solely measure pixel overlap between predictions and ground truth, cIDice introduces a geometry and connectivity focused approach. Its core idea revolves around analyzing the centerlines of tubular structures. From the ground truth and prediction masks, the centerlines of segmented structures are extracted using morphological skeletonization, which reduces tubular structures to their central representations.

$$\text{cIDice}(V_P, V_L) = 2 \times \frac{\text{Tp}_{\text{prec}}(S_P, V_L) \times \text{Ts}_{\text{sens}}(S_L, V_P)}{\text{Tp}_{\text{prec}}(S_P, V_L) + \text{Ts}_{\text{sens}}(S_L, V_P)}$$

with

$$\text{Tp}_{\text{prec}}(S_P, V_L) = \frac{|S_P \cap V_L|}{|S_P|} \quad ; \quad \text{Ts}_{\text{sens}}(S_L, V_P) = \frac{|S_L \cap V_P|}{|S_L|}$$

V_L represents the ground truth mask, V_P the segmentation prediction, S_L the ground truth skeleton, and S_P the predicted skeleton.

Once centerlines are extracted, cIDice assesses the overlap between the ground truth and predictions in two directions:

- T_{prec} (Precision): Measures the extent to which the prediction's centerlines are included within the ground truth.
- T_{sens} (Sensitivity): Measures the reverse inclusion, determining how well the ground truth's centerlines are represented in the prediction.

The cIDice metric combines these two directional overlaps to yield a global score. By focusing on centerlines, cIDice avoids penalizing minor segmentation errors in the thickness of structures, making it more suitable for tubular geometries. However, it has limitations when applied to the finest blood vessels, as we will demonstrate using an example from the STARE dataset.

Consider the two ground truth masks from the STARE dataset. Using one as the ground truth and the other as the prediction, we calculate their skeletons and use them to compute the cIDice score 3.1. The process reveals a key limitation: the connectivity of the masks is compromised. This can be observed by overlaying the prediction mask with the ground truth skeleton during the T_sens calculation. At this stage, visualization shows masks with pronounced gaps, indicating that the ground truth skeleton extends beyond the prediction mask, breaking connectivity. As a result, the precision score is adversely affected, leading to a low cIDice score of 0.48, only marginally higher than the Dice score of 0.45.

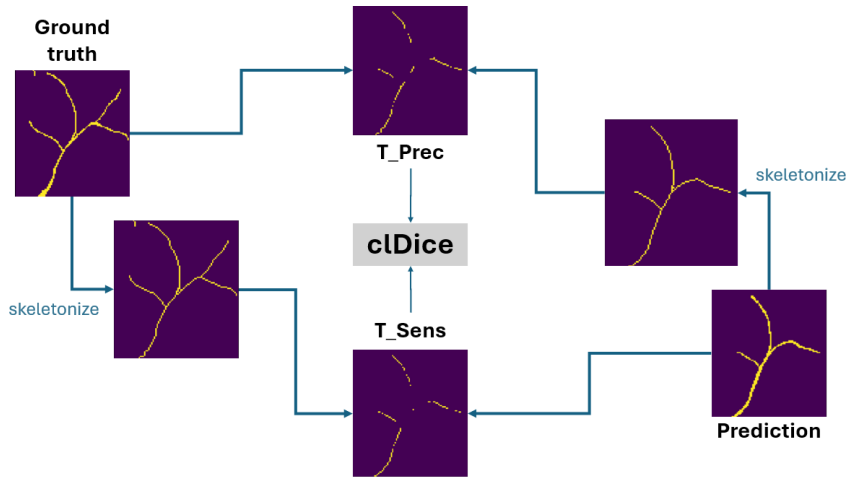


Figure 3.1: Diagram of the cIDice algorithm [Shit et al. \(2021\)](#), applied to the ground truths from the example in 2.2.

Despite its limitations, cIDice remains a highly valuable metric for tubular structure segmentation and serves as a solid baseline for future metric comparisons.

3.2 Our addition to cIDice: Smooth cIDice

Our enhancement to cIDice, which we call Smooth cIDice, builds on an earlier study conducted last semester. In that study, we explored smoothing masks to eliminate binary values and computed new metrics based on the smoothed masks. While the concept was intriguing, it lacked robustness, prompting us to shift focus after discovering cIDice. However, we retained the idea of introducing smoothing to create an uncertainty zone around blood vessels.

The final approach integrates this smoothing step into the cIDice algorithm. From the original

mask, we continue to calculate its morphological skeleton. Additionally, we compute a distance map of the mask using the [Distance map function](#) from the [Scipy library](#). The result is adjusted with an influence parameter that controls the degree of smoothing applied around the mask.

Returning to the example of the STARE dataset, we apply the Smooth cIDice algorithm [3.2](#). Visualization of the intermediate steps shows that connectivity is restored during the precision and sensitivity calculations. This improvement translates into higher metric scores, with a Smooth cIDice score of 0.60, a significant increase of 0.15 compared to the Dice score.

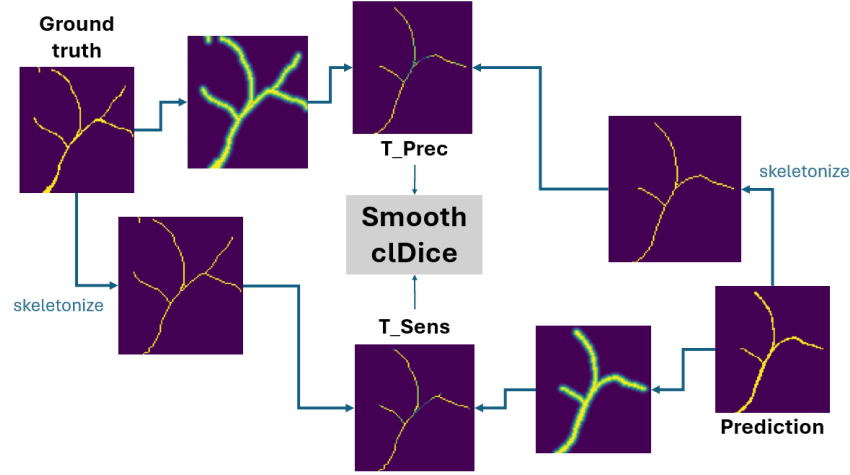


Figure 3.2: Diagram of the Smooth cIDice algorithm, applied to the ground truths from the example in Problematic [2.2](#).

We can then rewrite very simply the formula:

$$T_{prec}(S_P, D_L) = \frac{|S_P \cap D_L|}{|S_P|} \quad ; \quad T_{sens}(S_L, D_P) = \frac{|S_L \cap D_P|}{|S_L|}$$

D_L represents the smooth ground truth mask obtained after the distance map algorithm, D_P the smooth segmentation prediction, S_L the ground truth skeleton, and S_P the predicted skeleton.

3.3 Verification of metric's behavior

To ensure the Smooth cIDice metric is not inherently biased toward producing higher scores, we conducted controlled experiments.

- **Robustness to Small Uncertainties**

To test this, we progressively shifted one mask (prediction) relative to another (ground truth). For each shift, we calculated Dice, cIDice, and Smooth cIDice scores ([3.3](#)). The results show that both Dice and cIDice scores drop rapidly with small shifts, whereas Smooth cIDice is more tolerant to minor discrepancies before eventually decreasing when errors become significant.

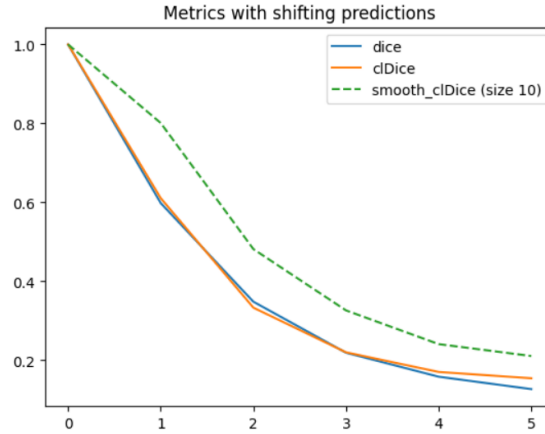


Figure 3.3: Evolution of the metrics (Dice, cLDice, smooth cLDice) with respect to the magnitude of the shift applied to the prediction (in pixels).

- **Sensitivity to Morphological Opening**

In the second experiment (3.4), we applied increasing levels of morphological opening to the prediction mask, gradually removing the finest vessels. This alteration significantly changes the topology, which should be penalized. Smooth cLDice exhibited stronger penalization for this error than the Dice score, and the same as the cLDice, confirming its sensitivity to topology disruptions.

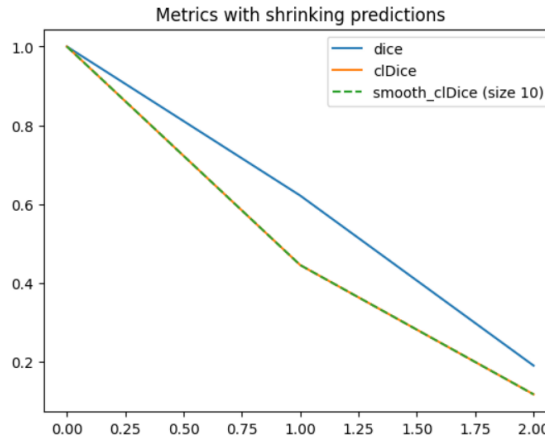


Figure 3.4: Evolution of the metrics (Dice, cLDice, smooth cLDice) with respect to the radius of the kernel used for the morphological opening applied to the prediction (in pixels).

3.4 Conclusion on the Smooth cIDice metric and future directions

These findings suggest that our metric holds promise. The small modification introduced to cIDice significantly impacts its behavior and aligns well with the challenges identified in our research problem.

Moving forward, we propose adapting Smooth cIDice as both a cost function to guide model training and an evaluation metric for segmentation tasks. This dual application could further validate its utility in enhancing segmentation performance for tubular structures.

Chapter 4

A gradient-based optimization compatible version of Smooth cLDice

4.1 Conditions to adapt the metric as a loss function

It may seem easy to adapt a metric as a loss function, but this is not always the case. Especially in our situation, where the metric involves several steps, each of which must meet specific conditions to be used as a loss function.

The primary condition for our metric is the differentiability of the function. This ensures that gradients can propagate during the backpropagation phase. To optimize our model by minimizing the loss function, gradients must pass through the loss function. If it does not allow gradient propagation, backpropagation becomes impossible, and our network cannot learn.

Looking at our metric, several parts of the algorithm require particular attention: the computation of skeletons for the masks and the smoothing of the masks.

4.2 The skeletonization algorithm

First, the skeletonization operation is well-documented, and various implementations are available ([Shit et al. \(2021\)](#), [Menten et al. \(2023\)](#), [Guzzi et al. \(2024\)](#)). Why so many implementations? Well, the original paper introducing cLDice also proposed a differentiable version to use as a loss function for our models. The issue is that the [skeletonization function](#) from the [scikit-image package](#), which computes the morphological skeleton of a mask, is not differentiable and thus cannot be used. To address this, a "soft skeleton" function utilizing pooling operations in PyTorch was proposed in the original 2021 paper, solving this problem.

However, as shown in this [table 4.1](#) from a study comparing various methods for obtaining a morphological skeleton [Menten et al. \(2023\)](#), not all methods are equal. The method introduced in the original paper has significant shortcomings. Specifically, it fails to preserve the topology of blood vessels, which is critical when the primary goal of our network is to maintain topology. Constructing an entire loss function to force the network to preserve topology while destroying it in the loss function's computations seems counterproductive.






Input	Non-differentiable skeletonization	Morphological skeletonization	Neural-network-based skeletonization	Our skeletonization
				
Compatible with gradient-based optimization	✗	✓	✓	✓
Topology-preserving	✓	✗	✗	✓
Domain-agnostic	✓	✓	✗	✓

Figure 4.1: Comparison of different skeletonization algorithms. [Menten et al. \(2023\)](#)

Nevertheless, we initially implemented the smooth cIDice using this method, as it provides acceptable results in 2D. However, more recent papers suggest new methods for extracting the morphological skeleton. This table 4.1 is from one such paper, which presents a function compatible with gradient propagation and capable of preserving blood vessel topology.

Although we haven't integrated this new skeletonization function into our smooth cIDice loss yet, we found an implementation of cIDice Loss that uses it and plan to include it in our initial experiments in part 5.

Regarding performance, the fastest algorithm is the scikit-image function, which is unsuitable for our loss function. This is followed by the iterative morphological operations version, while the topology-preserving, gradient-compatible function is far slower. Despite its excellent results, its slowness could significantly impact our model's training time.

4.2.1 The skeletonization algorithm based on iterative pooling functions

While this skeletonization algorithm is relatively straightforward to implement in [PyTorch](#), understanding it is essential 4.2. Starting with the original mask (either a prediction or ground truth), we first apply a min pooling operation to simulate morphological erosion with a 2x2 structuring element. The result is then subjected to max pooling, simulating dilation. Together, these steps simulate a morphological opening, which removes the thinnest blood vessels. By taking the difference, we perform a white top-hat operation, extracting the thin blood vessels removed during the opening. Applying a ReLU function eliminates negative values.

The output consists of single-pixel blood vessels removed during the opening, roughly corresponding to their skeleton due to their thickness. This process produces the result for the first iteration. Repeating these calculations, starting with the eroded mask, progressively refines the mask, reconstructing the entire skeleton by adding the outputs of each iteration.

The final result 4.3, obtained from a DRIVE dataset ground truth mask, is not perfect. There are gaps and regions thicker than a single pixel, but it suffices for now.

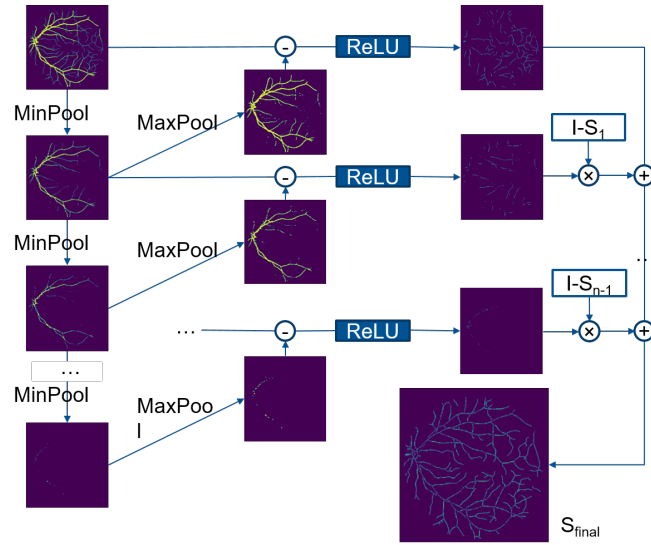


Figure 4.2: Algorithm for a differentiable skeletonization using pooling functions. [Shit et al. \(2021\)](#)



Figure 4.3: mask of the DRIVE dataset [Staal et al. \(2004\)](#) before application of the differentiable skeletonization algorithm.

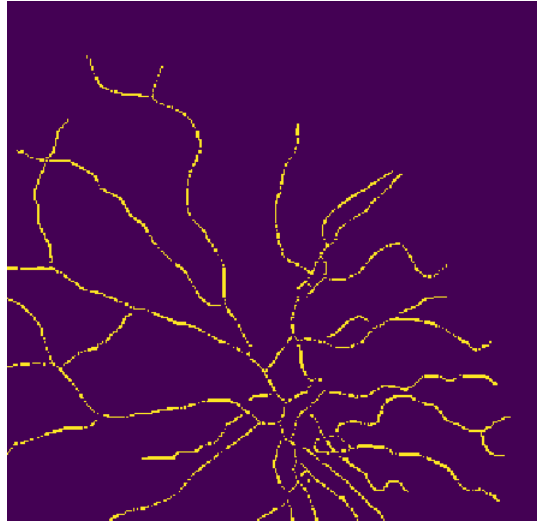


Figure 4.4: Result of the differentiable skeletonization algorithm [Shit et al. \(2021\)](#) on the mask.

4.3 Smoothing function, distance map algorithm

Now that we've chosen a skeletonization function, we can move on to mask smoothing. While I've briefly explained the principle in the chapter 3.2, this function suffers from the same problem as skeletonization: it uses a distance map calculation function which is not differentiable. Thus, we drew inspiration from the cIDice skeletonization algorithm to achieve a similar result

to the scipy function output using morphological operations (pooling operations in PyTorch).

We iterated multiple times to create an algorithm that closely replicates scipy function's behavior. Regarding performance, once again, the unusable scipy function is the fastest: 0.27 milliseconds, compared to our iterative algorithms, which take 2 and 7 milliseconds for the simple and advanced versions, respectively.

4.3.1 First iteration, simple algorithm

In the first version 4.5, for a smoothing factor n , we apply n iterations of the following steps:

- Perform dilation, similar to skeletonization, using max pooling.
- Compute the difference to extract the edges of the dilated vessels. These pixels correspond to those one pixel away from the original mask.
- Add these pixels to the distance mask, weighted inversely by their distance from the original mask.

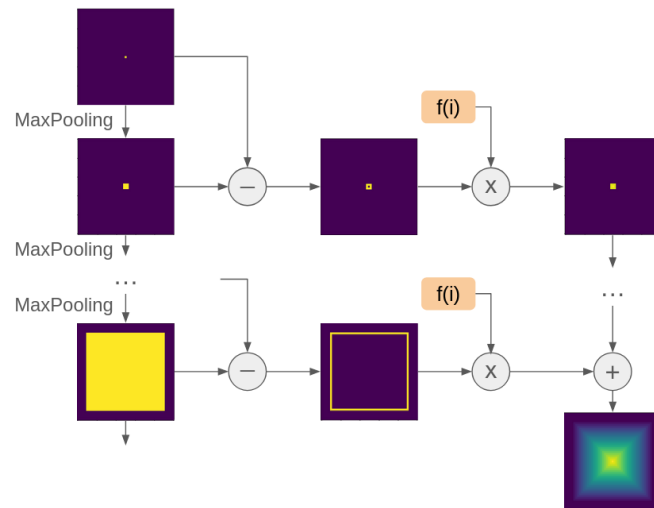


Figure 4.5: Basic algorithm of the smoothing function using the distance map

Applying this to a pixel results in a distance map that does not perfectly respect Euclidean distance to the center, as pooling operations use a square structuring element. For this reason, we propose another version of the smoothing function algorithm, more complex, but that gives better results.

4.3.2 Second iteration, complex algorithm

In the second iteration of the algorithm 4.6, which yields the best results we achieved, pooling is decomposed into two operations with different structuring elements: a 3x3 square and a cross. Additionally, using different factors for each direction improves the mask's approximation to a true Euclidean distance map compared to the first version.

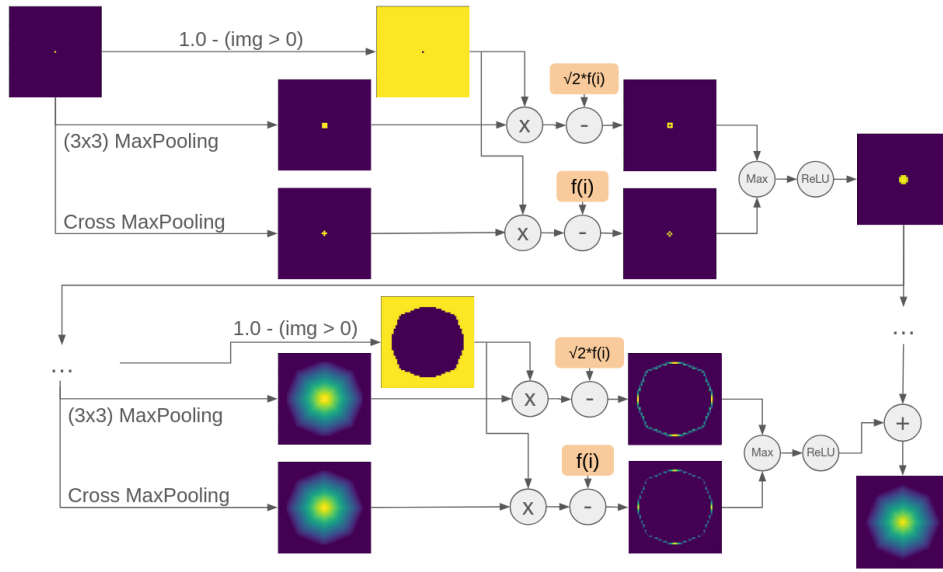


Figure 4.6: Advanced algorithm for the smoothing function using the distance map.

We still use the smoothing factor n , as the hyperparameter that let us decide the strength of the smoothing.

4.4 Limits of the loss function, practical application

The primary issue with this cost function is that it cannot independently guide model training effectively. When using this cost function alone, the results often lead to over-segmentations that are entirely unusable.

By analyzing the equations for the smooth cLDice and the original cLDice, it becomes evident that neither penalizes over-segmentation. As long as the skeleton of the over-segmented mask matches the "ideal" mask's skeleton, the network tends to favor over-segmentation. This is because over-segmentation minimizes topological errors, avoiding gaps in the prediction. Consequently, the model prioritizes maintaining the topology rather than producing accurate mask boundaries.

To counteract this effect, it is necessary to combine the smooth cLDice cost function with another cost function that prevents over-segmentation. Traditional loss functions, such as binary cross-entropy or Dice loss, are suitable for this purpose. These functions penalize over-segmentation by emphasizing the agreement between predicted and ground truth masks in terms of area.

The approach involves calculating both loss functions and combining their results by adding them with a weighting factor. The choice of this weighting factor determines the balance between emphasizing the preservation of the network's topology and the accurate segmentation of the area. This compromise allows the model to produce predictions that better align with the desired segmentation quality.

Chapter 5

Experience: loss functions comparison

We introduced an implementation of our cost function, made several assumptions, and now aim to experimentally evaluate it. To test these cost functions and their influence on model training, we designed a straightforward protocol:

5.1 Experience protocol

5.1.1 Datasets Selection

We chose several datasets for the experiments. For the initial trial, we relied on familiar datasets that we had previously worked with, ensuring sufficient diversity. Specifically, we used three databases: DRIVE, CHASE, and HRF. These datasets consist of retinal fundus images featuring blood vessels:

- **DRIVE:** [Staal et al. \(2004\)](#)
A reference dataset for blood vessel segmentation. Although it is of relatively low quality, it is essential for benchmarking due to its widespread use and strong baseline characteristics.
- **CHASE:** [Fraz et al. \(2012\)](#)
Offers moderate quality, higher than DRIVE, and serves as a good intermediary. This dataset also provides two ground truths for each image.
- **HRF:** [Budai et al. \(2013\)](#)
A high-quality dataset providing detailed vessel segmentation.

This diversity allows us to test our metrics and cost functions on data of varying resolutions, which is particularly relevant given the challenges discussed earlier, such as partial volume effects and the resolution of fine blood vessels.

5.1.2 Metric Selection

Having selected the datasets, the next step was identifying metrics to evaluate the training performance. Alongside standard metrics (despite their limitations), we included the following: cIDice Score, smooth cIDice and betti error metrics.

Betti error metrics

The Betti error metrics leverage Betti numbers, which characterize topological features. For instance:

- Betti number 0 represents the number of connected components in the surface.
- Betti number 1 corresponds to the number of independent closed curves, equating to the number of holes in the surface in 2D.

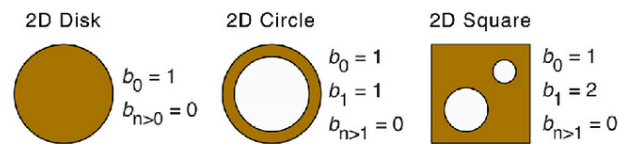


Figure 5.1: Illustration of Betti numbers in 2 dimensions [Stenseke \(2021\)](#)

Using these numbers, we compute similarity scores between the ground truth and predictions. The closer the Betti numbers of the prediction are to the ground truth, the better the prediction respects the topology of the blood vessels. This results in a Betti error approaching zero.

Unlike other metrics, Betti errors are more reliable. Ground truth imperfections, discussed earlier, do not heavily impact this metric, as the topology of ground truth images generally reflects the true topology of the blood vessels. This makes Betti errors particularly valuable for training analysis.

5.1.3 Cost Functions Selection

As previously discussed, our cost function cannot be used in isolation. Thus, we introduced a ratio hyperparameter to balance it with other loss functions during training. Additionally, the smoothing strength hyperparameter allows further customization of the cost function, yielding various configurations.

Given the large dataset size and data augmentation requirements, and to avoid excessive computational load, we limited the number of cost functions to test. The selected functions include:

- **Weighted Binary Cross-Entropy Loss (BCE Loss):** A standard loss function we have previously employed.
- **Dice Loss:** Another traditional loss function, based on the dice score metric.
- **cIDice Loss:** Using a skeletonization method from the 2024 study, making it a noteworthy addition for comparison.

- **Smooth cIDice Variants:** Six versions of our cost function, derived by varying hyperparameters:
 - Smoothing strength: 0, 2 and 5
 - Ratio between the complementary loss and Smooth cIDice: 0.5 and 0.8

In total, we tested nine cost functions across three datasets, resulting in 27 training experiments. This experimental setup ensures robust evaluation of our cost function and its impact on model performance.

5.1.4 Model

For this application, we selected the U-Net architecture [Ronneberger et al. \(2015\)](#), a widely used convolutional neural network (CNN) in biomedical image segmentation.

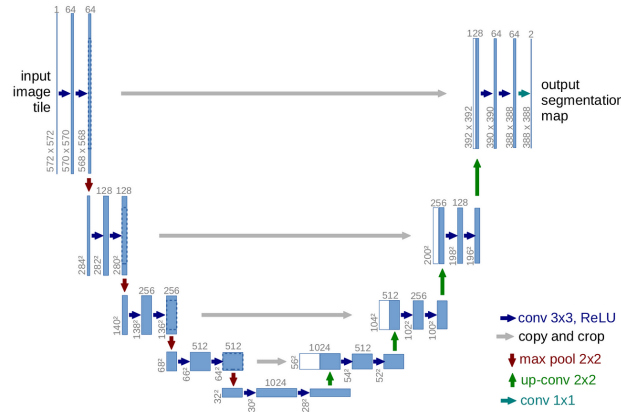


Figure 5.2: The U-Net model architecture.

The U-Net features a symmetrical encoder-decoder structure [5.2](#): the encoder progressively reduces spatial dimensions while capturing contextual information, and the decoder upsamples the compressed features to produce a segmentation mask. Skip connections between corresponding encoder and decoder layers preserve spatial details, which is critical for accurately segmenting thin and intricate structures like blood vessels.

The U-Net's ability to balance local and global information makes it well-suited for this task, as it ensures precise segmentation of fine vessel networks while maintaining coherence across the entire retinal image.

5.1.5 Training

For each training session, we employ a simple train/validation/test split to evaluate the performance of our model. The dataset is divided into three distinct subsets: a training set for model learning, a validation set for monitoring the training process and guiding decisions such as early stopping, and a test set for the final performance evaluation. This setup ensures that

the model is optimized during training while also providing an unbiased assessment of its generalization capabilities on unseen data.

While k-fold cross-validation would have been the ideal approach for this study due to its ability to provide a more robust and generalized evaluation of the model's performance, implementing it was not feasible within the available timeframe. K-fold cross-validation involves splitting the dataset into multiple folds and training/testing the model iteratively on different partitions, which significantly increases the computational cost and training time. However, k-fold cross-validation will be implemented in future experiments to provide a more comprehensive analysis. The results of these additional evaluations will be analyzed and discussed in a forthcoming article, allowing us to further validate our findings and strengthen the conclusions of this study.

We then export the predictions for the test set of each training, saving a file for each prediction containing all relevant information about the training process, the loss function, the model, and the dataset used. Crucially, we also record the results of the evaluation metrics for each prediction, which will be of primary interest in subsequent analyses.

5.2 Results analysis

5.2.1 Global metrics results

The evaluation of the loss functions performances was conducted through a detailed comparison of the metrics 5.3. For conventional metrics, such as specificity, false positive rate or positive predictive value, our custom smooth cIDice loss variants slightly underperformed compared to more traditional approaches. This trend wasn't consistent across other pixel-wise metrics, as our losses have some decent results in accuracy and Dice score.

However, when focusing on topology-oriented metrics, such as cIDice, the smooth cIDice loss clearly outperformed. Notably, the best results were consistently observed for configurations with a loss ratio of 0.5 compared to 0.8 in the mixed loss function. This indicates that assigning more weight to the smooth cIDice component enhances the preservation of vessel topology, leading to superior cIDice scores. A similar trend was observed with Betti error metrics, where lower mean errors were achieved using the smooth cIDice loss with a 0.5 ratio, despite the variance remaining relatively high. These findings underscore the capacity of the smooth cIDice loss to prioritize topological accuracy effectively.

Given the large number of metrics analyzed, it can be hard to analyse the results graphically. However, a comprehensive table containing the average results for each metric across all experiments, with the best values highlighted, is provided for reference 5.1.

5.2.2 Results by database

The experiments also investigated performance variations across datasets. While overall statistics aggregated predictions from all datasets, analyzing individual datasets revealed subtle differences 5.4.

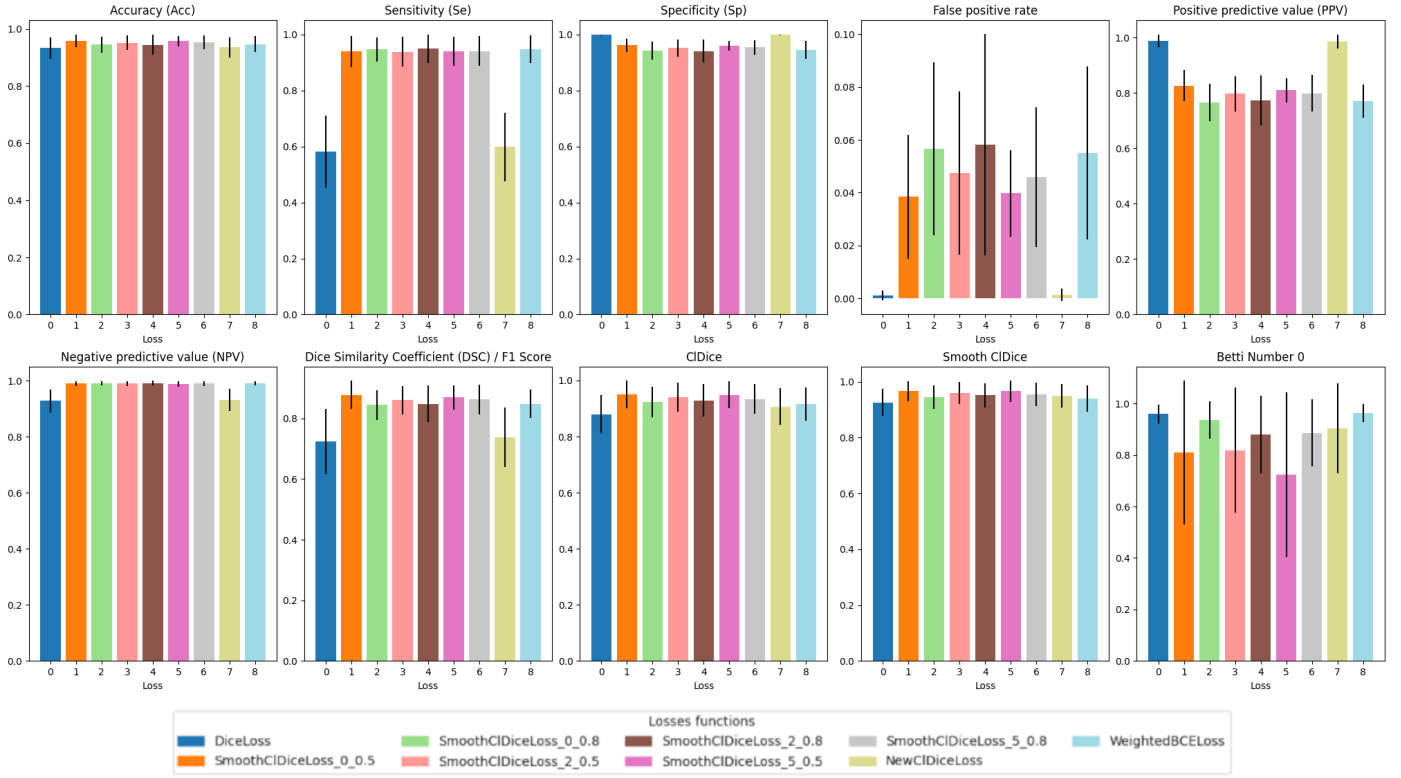
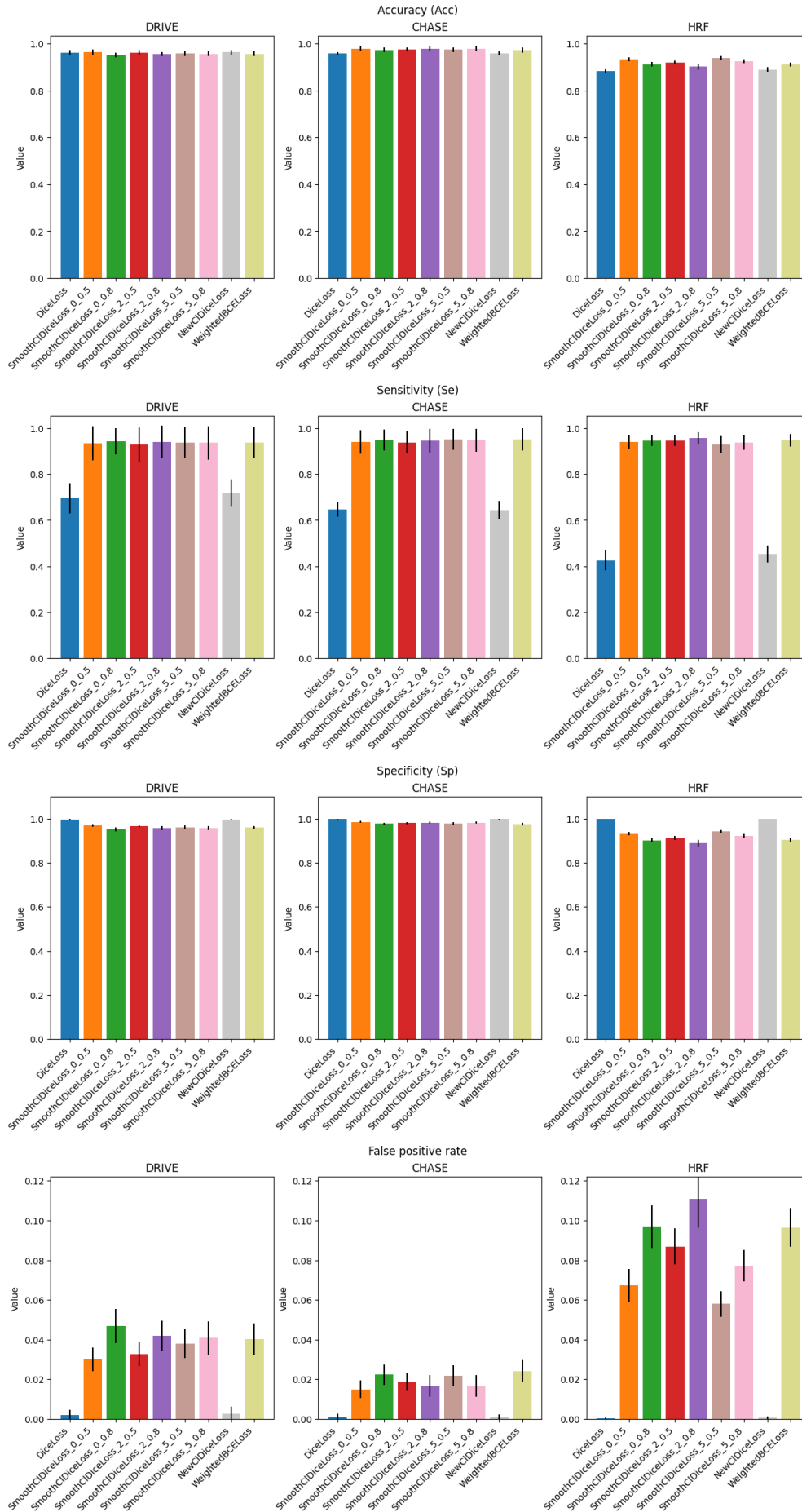


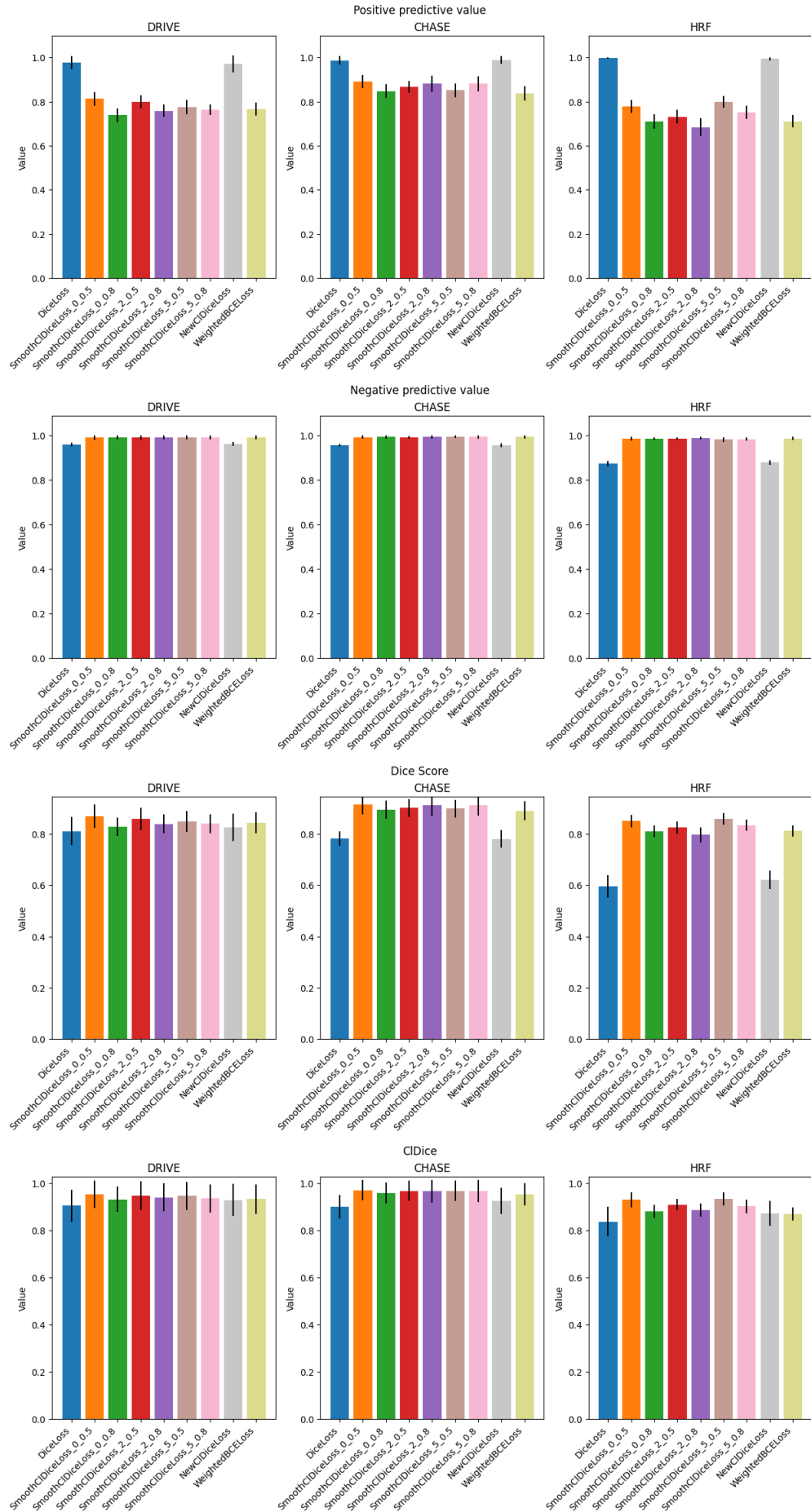
Figure 5.3: Mean and variance of the results according to the loss function, for all metrics.

	Accuracy	Sensitivity	Specificity	False positive rate	PPV	NPV	Dice Score	CIDice	Smooth CIDice	Betti Number 0
DiceLoss	0.933	0.583	0.999	0.001	0.988	0.928	0.724	0.879	0.926	0.959
SmoothCIDiceLoss_0_0.5	0.959	0.939	0.962	0.038	0.826	0.989	0.878	0.95	0.966	0.81
SmoothCIDiceLoss_0_0.8	0.945	0.947	0.943	0.057	0.764	0.99	0.844	0.922	0.945	0.935
SmoothCIDiceLoss_2_0.5	0.952	0.939	0.953	0.047	0.798	0.989	0.861	0.941	0.959	0.819
SmoothCIDiceLoss_2_0.8	0.945	0.949	0.942	0.058	0.773	0.991	0.848	0.929	0.951	0.879
SmoothCIDiceLoss_5_0.5	0.957	0.94	0.96	0.04	0.809	0.989	0.869	0.949	0.966	0.725
SmoothCIDiceLoss_5_0.8	0.953	0.941	0.954	0.046	0.799	0.989	0.862	0.934	0.954	0.886
NewCIDiceLoss	0.935	0.599	0.999	0.001	0.986	0.93	0.737	0.907	0.95	0.904
WeightedBCELoss	0.947	0.947	0.945	0.055	0.77	0.99	0.848	0.916	0.939	0.963

Table 5.1: Global results table, best column result is in bold.

For instance, the HRF dataset consistently yielded lower metric values. Upon examining the predictions, it appears this could be attributed to the high resolution of HRF images. The preprocessing steps, constrained by the fixed input size of the model, may degrade the fine details of smaller vessels, thus affecting performance.





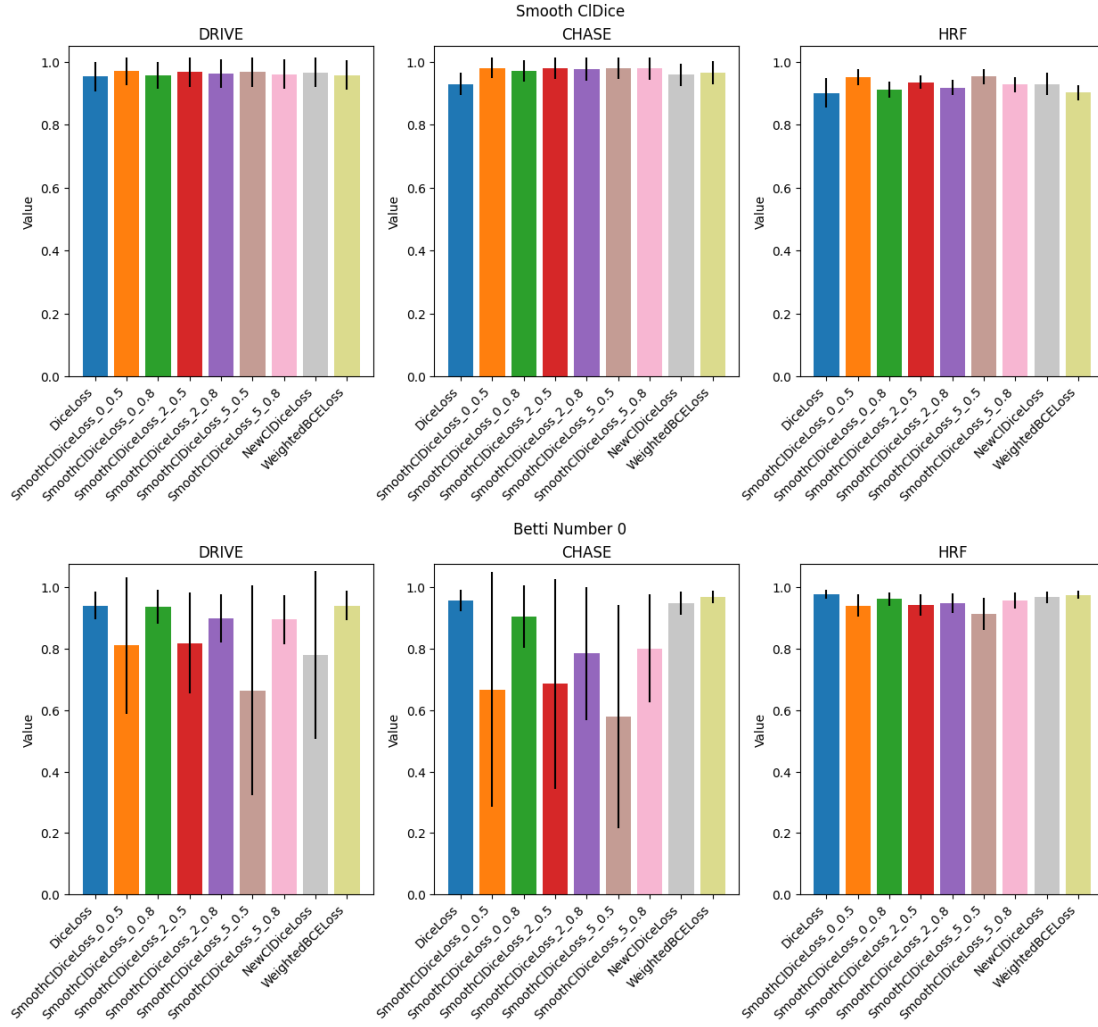


Figure 5.4: Mean and variance of the results according to the loss function, for all metrics, by database.

5.2.3 Ranking of loss functions

To better understand the performance of each loss function, we ranked them based on each metric 5.5. While this detailed ranking provides valuable insights, it is also lengthy and complex. To simplify, we computed an overall ranking by combining all metrics 5.6. Although this aggregate ranking is not entirely representative of the true quality of each loss function, given that it assigns equal weight to all metrics, some of which are less relevant, it still offers meaningful observations.

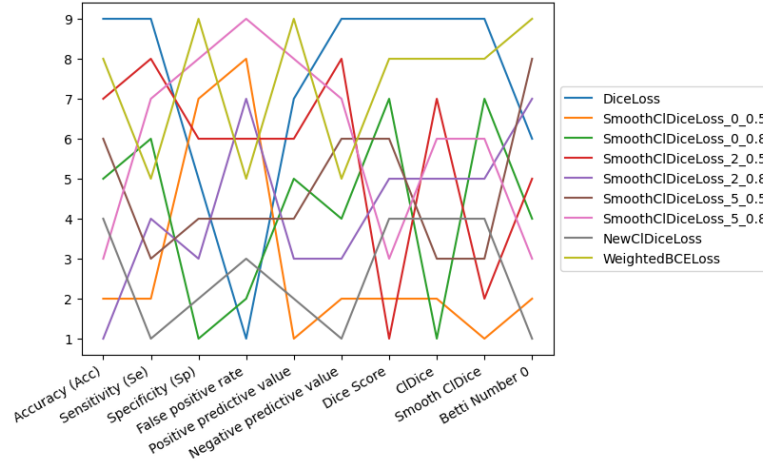


Figure 5.5: Ranking of loss functions for each metric.

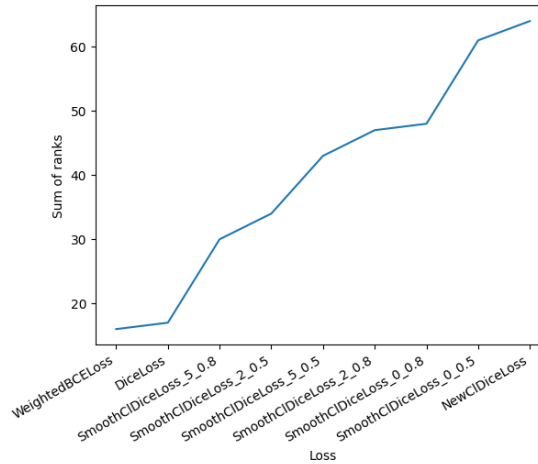


Figure 5.6: Global ranking of loss functions, with the sum of their rank on each metric.

The results highlight that all training runs using the smooth cIDice loss ranked highly. The only loss function outperforming it in the general ranking was a cIDice loss variant employing a recent and reportedly more effective skeletonization function. This observation underscores the critical impact of choosing an appropriate skeletonization approach when designing topology-focused loss functions.

Chapter 6

Future work

This study represents a significant step forward in optimizing loss functions for vascular segmentation, but several critical areas remain to be addressed to refine and expand the scope of the project. These areas are outlined below, providing a roadmap for continued investigation:

- **Analyzing the Influence of Skeletonization Methods**

Future work will involve systematically evaluating the impact of using different skeletonization techniques. This includes assessing the topological accuracy of resulting masks and understanding how changes in skeletonization affect the performance of the loss function.

- **Improving Smoothed Masks and Distance Maps**

Efforts should focus on developing a robust method for generating smoothed masks and associated distance maps. This includes a more detailed analysis of the smooth factor hyperparameter, determining its optimal value, and studying its influence on segmentation performance. A deeper exploration into different smoothing techniques and their effect on distance maps could provide further insights into achieving a balance between topological fidelity and segmentation accuracy.

- **Exploring the Combined Loss Function**

The combination of the smooth cLDice loss with another auxiliary loss function remains an area requiring extensive experimentation. Investigating the choice of auxiliary loss, the hyperparameter governing their ratio, and even dynamic strategies for varying this ratio during training could lead to significant improvements. For instance, transitioning from a fixed ratio to a time-varying ratio might enable the network to balance topology and pixel accuracy adaptively. The current framework appears ready for these experiments, requiring primarily the execution of training runs with various parameter configurations and subsequent analysis of results.

- **Finalizing and Validating the Loss Function**

To conclude this study, it is imperative to finalize the implementation of the smooth cLDice loss by selecting the optimal skeletonization and mask-smoothing functions and fixing hyperparameter values. A comprehensive evaluation of this finalized loss function, through full training runs, will allow for a more precise comparison against benchmarks and provide conclusive evidence of its efficacy.

- **Implementing Additional Loss Functions**

The integration and validation of other loss functions, such as the cICE Loss [Acebes et al.](#)

(2024) and cbDice Loss [Shi et al. \(2024\)](#), remain pending. Including these in the experimental framework will enable broader comparisons and insights into how different loss functions perform under varied conditions and datasets.

- **Expanding Dataset Diversity**

Expanding the dataset scope by incorporating additional 2D datasets, such as ROSE [Ma et al. \(2020\)](#), and extending the study to 3D datasets is crucial. The inclusion of diverse and higher-dimensional data will test the robustness of the proposed methods across different imaging modalities and resolutions.

6.1 Supervision of the project

Given the personal and professional significance of this project, as well as its incomplete state and the opportunities for new ideas, we decided to ensure its continuity by transitioning to supervisory roles. To achieve this, we have onboarded a new RDI (Research and Development in Innovation) student who will take over the project and carry it forward. This arrangement allows us to maintain a connection with the research, provide guidance to the student, and regularly monitor the project's progress. By maintaining periodic meetings, we aim to ensure that the project's objectives remain aligned with its original vision and that its development proceeds systematically.

In addition to the RDI student, there is the possibility of involving an intern who will work on a closely related topic. This parallel effort could provide valuable opportunities for collaboration, data sharing, and cross-validation of results.

6.2 Metric coherence verification

A potential avenue for the future student involves exploring a topic that we only briefly touched upon: assessing the relevance and effectiveness of evaluation metrics. Unlike evaluating models or loss functions, scoring metrics themselves poses unique challenges, as metrics typically serve as the benchmark for assessing performance. In this case, we aim to rank these evaluation indicators based on their alignment with human perception, particularly in the context of medical applications.

The most promising approach identified involves comparing metric outcomes with human preferences, specifically those of domain experts like physicians or experienced annotators. To achieve this, we envisioned a platform capable of collecting user feedback on segmentation predictions. The primary goal would be to gather extensive data on preferences between pairs of predictions and correlate this feedback with the corresponding metric scores. While assigning an exact numerical score to a prediction is complex, a simpler task involves determining which of two predictions is preferred based on visual inspection.

For example [6.1](#), users might be presented with an image, and predictions from two different trainings, and asked which prediction better captures the blood vessels. In cases where one prediction avoids false negatives (e.g., vessel gaps) while the other suffers from such errors, users would likely prefer the former, even if it lacks perfect edge precision.

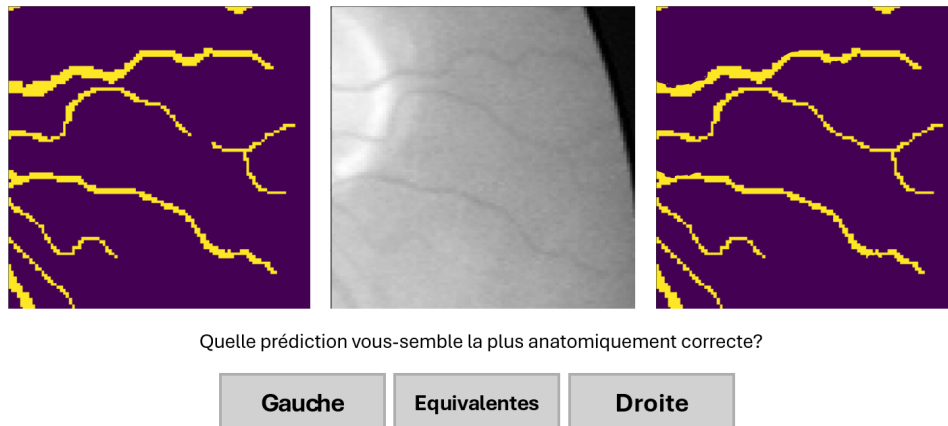


Figure 6.1: Example of the user interface for the platform

Leveraging the metadata generated in this research, such as metric results and associated predictions, the platform could facilitate statistical analysis comparing user preferences to metric scores. This comparison would highlight the consistency (or not) between metrics and human judgment. By confronting metric results with professional opinions, the project would provide valuable insights into the validity and reliability of metrics for specific tasks. Furthermore, it opens the possibility of defining "metrics for metrics", that are quantitative measures of how well a metric aligns with human perception. This could serve as an objective framework for assessing the utility of metrics like smooth cDice in comparison to others.

Chapter 7

Conclusion

This research addressed the challenges of accurately segmenting vascular structures, particularly thin and intricate blood vessels, which are often inadequately captured by traditional pixel-wise metrics. Our primary goal was to develop a method that better aligns with the topological and structural properties of these networks.

To this end, we introduced and refined the smooth cLDice metric, an adaptation of the cLDice metric aimed at improving robustness through mask smoothing based on distance map computing. Building on this, we proposed the smooth cLDice Loss, a hybrid loss function that combines structural and pixel-wise evaluation. The integration of a tunable weighting parameter allowed us to balance the emphasis on topology and pixel precision, providing flexibility for different segmentation challenges.

We conducted extensive experiments using public vascular datasets, and results revealed that while our smooth cLDice Loss did not particularly outperform all traditional losses on pixel-wise metrics like specificity, it achieved significantly better results on topology-focused metrics such as cLDice. Furthermore, variations in the loss function's weighting parameter demonstrated that increasing the emphasis on the smooth cLDice component consistently improved performance on topological metrics.

Overall, this study provides a solid foundation for topology-aware segmentation and underscores the value of considering structural metrics in medical imaging. Future work will aim to enhance the smooth cLDice Loss, extend the evaluation to additional datasets, including 3D vascular imaging, and design and analyse the results of a metric coherence verification platform.

I am also thrilled to share that our abstract has been accepted for presentation at the IABM 2025 conference in Nice, France. This provides an incredible opportunity to showcase our work on vascular segmentation and the development of the smooth cLDice metric to a broader scientific audience. Presenting at such an event will allow us to gather valuable feedback, engage in discussions with experts in the field, and further highlight the significance of our contributions to medical image analysis.

I'd like to finish this report by warmly thanking Elodie for her support and guidance throughout the year, this project would never have seen the light of day without her involvement. I really enjoyed working on it and it's mainly thanks to her.

Chapter 8

Bibliography

Acebes, C., Moustafa, A. H., Camara, O., and Galdran, A. (2024). The centerline-cross entropy loss for vessel-like structure segmentation: Better topology consistency without sacrificing accuracy. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 710–720. Springer. (pages 6 and 29)

Budai, A., Bock, R., Maier, A., Hornegger, J., and Michelson, G. (2013). Robust vessel segmentation in fundus images. *International journal of biomedical imaging*, 2013(1):154860. (page 20)

Fraz, M. M., Remagnino, P., Hoppe, A., Uyyanonvara, B., Rudnicka, A. R., Owen, C. G., and Barman, S. A. (2012). An ensemble classification-based approach applied to retinal blood vessel segmentation. *IEEE Transactions on Biomedical Engineering*, 59(9):2538–2548. (page 20)

Givisiez, R. (2025). Blood-vessel-segmentation. Accessed: 2025-01-18. (page 5)

Guzzi, L., Zuluaga, M. A., Lareyre, F., Di Lorenzo, G., Goffart, S., Chierici, A., Raffort, J., and Delingette, H. (2024). Differentiable soft morphological filters for medical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 177–187. Springer. (page 15)

Hoover, A. and Goldbaum, M. (2000). STARE: STructured Analysis of the Retina. <http://cecas.clemson.edu/~ahoover/stare/>. Accessed: 2024-10-10. (page 8)

Ma, Y., Hao, H., Xie, J., Fu, H., Zhang, J., Yang, J., Wang, Z., Liu, J., Zheng, Y., and Zhao, Y. (2020). Rose: a retinal oct-angiography vessel segmentation dataset and new model. *IEEE transactions on medical imaging*, 40(3):928–939. (page 30)

Menten, M. J., Paetzold, J. C., Zimmer, V. A., Shit, S., Ezhov, I., Holland, R., Probst, M., Schnabel, J. A., and Rueckert, D. (2023). A skeletonization algorithm for gradient-based optimization. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 21394–21403. (pages 15 and 16)

Mostapha, M. (2014). A novel diffusion tensor imaging-based computer-aided diagnostic system for early diagnosis of autism. (page 9)

Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pages 234–241. Springer. (page 22)

- Shi, P., Hu, J., Yang, Y., Gao, Z., Liu, W., and Ma, T. (2024). Centerline boundary dice loss for vascular segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 46–56. Springer. (pages 6 and 30)
- Shit, S., Paetzold, J. C., Sekuboyina, A., Ezhov, I., Unger, A., Zhylka, A., Pluim, J. P., Bauer, U., and Menze, B. H. (2021). cldice-a novel topology-preserving loss function for tubular structure segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 16560–16569. (pages 10, 11, 15, and 17)
- Staal, J., Abramoff, M., Niemeijer, M., Viergever, M., and van Ginneken, B. (2004). Ridge-based vessel segmentation in color images of the retina. *IEEE Transactions on Medical Imaging*, 23(4):501–509. (pages 17 and 20)
- Stenseke, J. (2021). Persistent homology and the shape of evolutionary games. *Journal of Theoretical Biology*, 531:110903. (page 21)
- Stucki, N., Paetzold, J. C., Shit, S., Menze, B., and Bauer, U. (2023). Topologically faithful image segmentation via induced matching of persistence barcodes. In *International Conference on Machine Learning*, pages 32698–32727. PMLR. (page 6)