A skeletonization algorithm for gradient-based optimization – supplementary material

Martin J. Menten[1,2] Johannes C. Paetzold[1,2] Veronika A. Zimmer[1] Suprosanna Shit[1]
Ivan Ezhov[1] Robbie Holland[2] Monika Probst[1] Julia A. Schnabel[1] Daniel Rueckert[1,2]
[1]Technical University of Munich [2]Imperial College London

Used datasets

DRIVE The widely used DRIVE dataset consists of 40 two-dimensional retinal color fundus photographs and matching annotations of the visible blood vessels [9]. We normalize the images to an intensity range between 0 and 1 and crop them to a size of 512×512 pixel. Then, we divide the dataset into training, validation and testing splits with a ratio of 60% to 20% to 20%.

VesSAP The VesSAP dataset contains 24 three-dimensional light-sheet microscopy images of murine brains after tissue clearing, staining, and labeling of the vascular network. It has been made publicly available and has been extensively described by Todorov *et al*. [10]. We split the $500 \times 500 \times 50$ voxel large images into non-overlapping patches of size $50 \times 50 \times 50$. We remove the patches that only contain background. Finally, we split the remaining ones into a training, validation and testing partition with a ratio of 80% to 10% to 10%, while ensuring a subject-wise split.

Mandible The mandible dataset consists of 34 matched CT and MR images of the lower head and neck. In all images the mandible bone was outlined by a clinical expert. We resample all images to a resolution of $0.25 \times 0.25 \times 0.25$ cm^3 and subsequently remove all smaller cavities of the segmentation mask by alternatingly applying dilation and erosion operations. For the benchmarking experiments of the skeletonization algorithm we exclusively use the CT images (cf. Section 4.1 of the main paper). For the multimodal registration workflow that incorporates our skeletonization module we use the matched image pairs (cf. Section 4.3 of the main paper).

Neural network architecture and training

This work uses neural networks either for explicit skeletonization (cf. Section 4.1 of the main paper) or for vessel segmentation (cf. Section 4.2 of the main paper). In the first case networks are provided with a binary mask and asked to provide the ground truth skeleton while being evaluated using the Dice loss. In the second case the neural network is provided images from either the DRIVE or VesSAP dataset and trained to output a blood vessel segmentation map. Hereby, we use the topology-preserving clDice loss in combination with various skeletonization algorithms [8].

The two neural-network-based skeletonization methods are implemented according to the works and accompanying public software code by Panichev *et al*. [6] and Nguyen [5], respectively. In order to facilitate processing of volumetric images we replace all two-dimensional operations, such as convolutions, pooling and normalization network layers with their three-dimensional equivalents.

The segmentation neural networks follow a basic U-Net architecture with four downsampling and four upsampling blocks with skip connections [7]. Each block consists of two sequences of convolutional layer (either two- or three-dimensional convolutions, kernel size: 3, same padding), instance normalization layer and leaky-ReLU non-linearity (slope: 0.01). Downsampling is achieved by using a stride of 2 in the second convolutional layer of each block. At each downsampling step, the number of feature maps is also doubled and copied to the skip connection. Upsampling is achieved via a transposed convolution with a kernel size of 2 and stride of 2. After upsampling, the respective skip connection is concatenated with the main feature map. Network weights are optimized using the ADAM optimizer with a learning rate of 10^{-4} [3]. In experiments using the DRIVE dataset, networks are trained with a batch size of 2 for 1,000 epochs. We apply random shifts ($\pm 10\%$) and rotations ($\pm 45°$) as data augmentation. For the VesSAP dataset, networks are trained with a batch size of 16 for 200 epochs. In each case, we pick the best performing network based on the validation dataset before reporting the results on the test dataset. All experiments are repeated three times using different random seeds.

Additional qualitative skeletonization results

Figure 1 presents additional representative results of applying the five skeletonization algorithms to the three datasets (cf. Section 4.1 of the main paper).

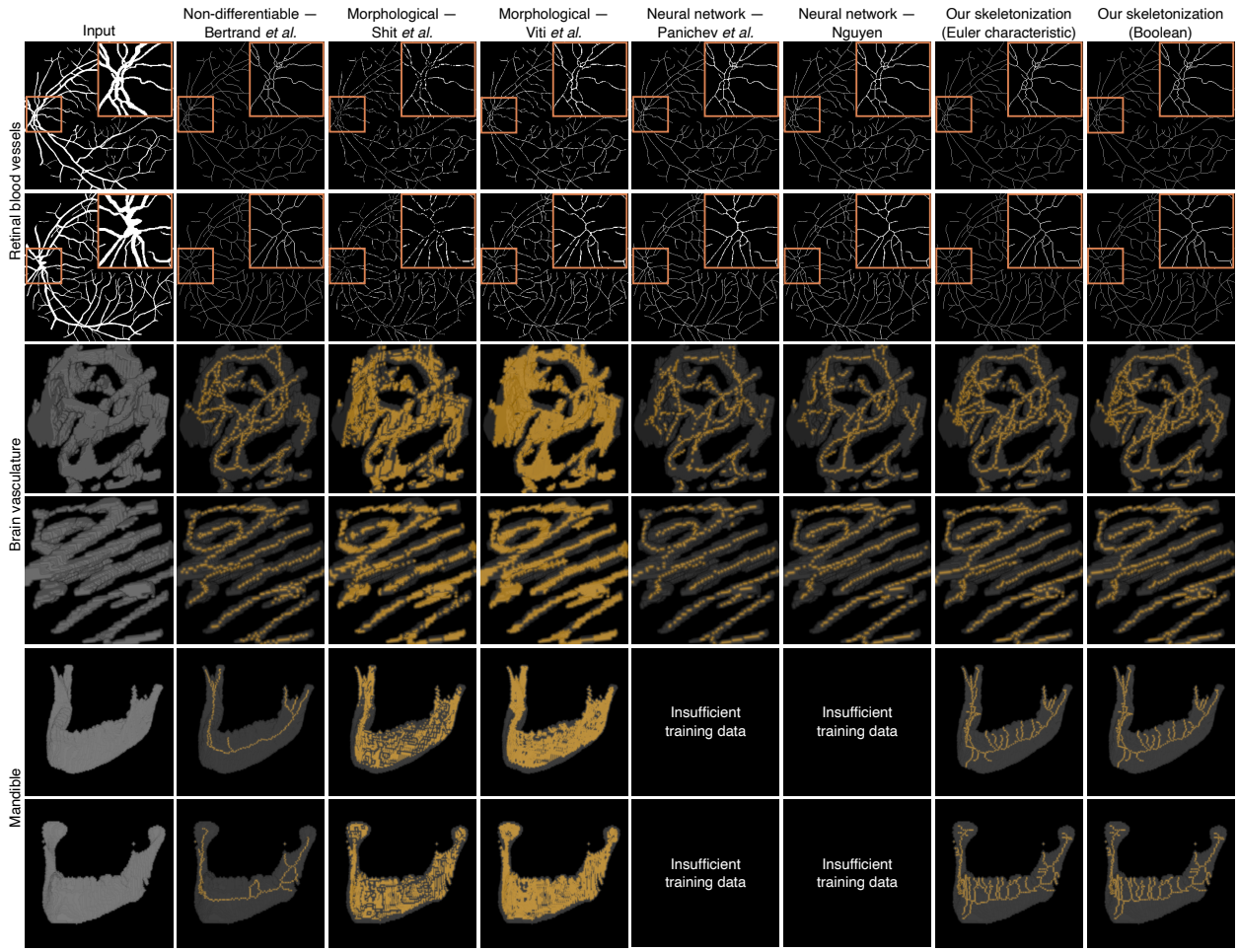

Figure 1. Additional results of applying the seven tested skeletonization algorithm to representative samples of three diverse datasets. Of the six algorithms that are compatible with gradient-based optimization, only our two methods are able to extract a thin, topology-preserving skeleton, similar to the one obtained using the non-differential baseline.

Experiments using the SkelNetOn dataset

The SkelNetOn dataset was published in the scope of the Deep Learning for Geometric Shape Understanding workshop held in conjunction with the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition [2]. It consists of binary images depicting a range of stylized objects and their corresponding skeletons. Compared to the complex topologies of biological structures, the dataset exclusively features closed two-dimensional surfaces.

We repeat the same benchmarking experiments as described in Section 4.1 of the main paper using the SkelNetOn dataset. At the time of our study the challenge's public leaderboard had been taken offline, so that we could not use the official test split to benchmark our skeletonization algorithms. Instead, we split the training dataset into a training, validation and testing partition. We observe the same characteristic behavior of all skeletonization algorithms as in the experiments with the other three datasets (see Figure 2). Both morphological baseline algorithms introduce breaks along the skeleton and in some cases omit large parts of the medial axis. The neural-network-based solutions also alter the topology of the object, whereas our skeletonization algorithms result in a topologically correct, thin skeleton. This is also reflected in the quantitative measurements reported in Table 1.

Effect of Boltzmann temperature on learning with the differentiable skeletonization module

The entropy of the stochastic discretization can be controlled by varying either the scale of the noise β or the Boltzmann temperature τ (cf. Equation 11 of the main paper). We have also conducted the simple experiment presented in Figure 6 of the main paper while varying τ instead of β. Hereby, an input tensor is initialized with random values and passed through our skeletonization module. Using backpropagation, the tensor's values are learned so that its

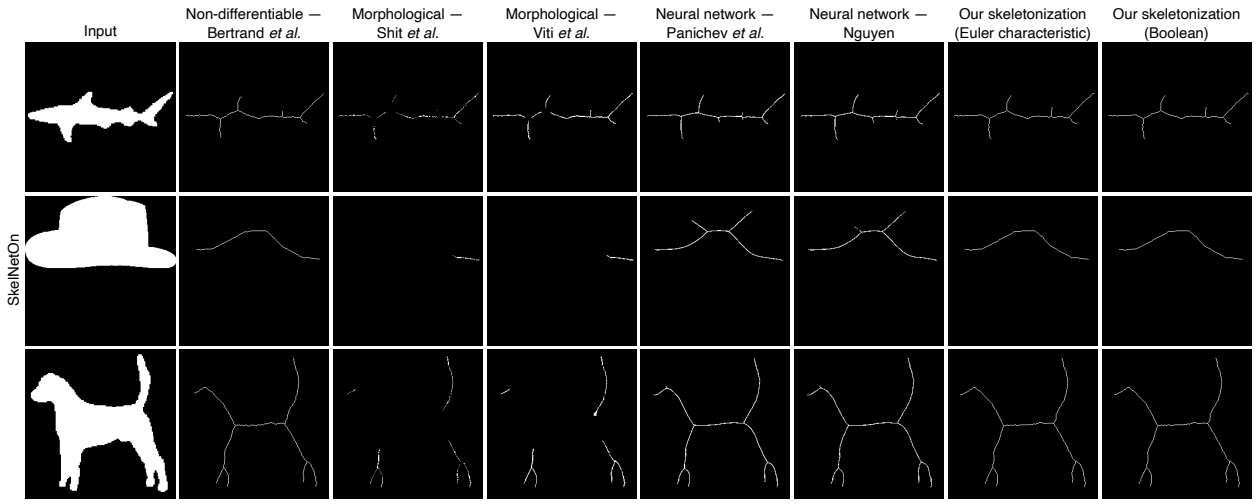

Figure 2. Qualitative results of applying the seven tested skeletonization algorithm to representative samples of SkelNetOn dataset.

Table 1. Quantitative comparison of the topological accuracy of seven skeletonization algorithms on the SkelNetOn dataset.

Dataset	Skeletonization algorithm	# points	β_0 error	β_1 error	β_2 error	Run time [ms]
SkelNetOn	Non-differentiable – Bertrand *et al.* [1]	355±206	0±0	0±0	-	-
	Morphological – Shit *et al.* [8]	158±150	40±30	0±1	-	23±2
	Morphological – Viti *et al.* [11]	355±295	5±5	0±1	-	26±2
	Neural network – Panichev *et al.* [6]	524±247	2±2	0±1	-	30±1
	Neural network – Nguyen [5]	494±236	2±3	0±1	-	160±2
	Ours – Euler characteristic	406±241	0±0	0±0	-	189±3
	Ours – Boolean	406±241	0±0	0±0	-	948±5

ultimate output resembles that of the ground truth skeleton. Analogously to our our results with varying noise scales (cf. Figure 7 of the main paper), we find that both a too low and too high Boltzmann temperature inhibit efficient learning with our skeletonization module (see Figure 3). Empirically, we find that it suffices to tune either the noise scale or Boltzmann temperature and proceed to tune β throughout all other presented experiments.

Vessel segmentation in the DRIVE dataset

As described above and in Section 4.2 of the main paper, we integrated our skeletonization modules with a neural network that learns to segment blood vessels in either the VesSAP or DRIVE dataset. The results on the two-dimensional DRIVE dataset are shown in Table 2. Similar to the results for the VesSAP dataset (see Table 2 of the main paper), we find that using the clDice loss instead of a vanilla Dice loss slightly improves the topological agreement between prediction and ground truth as indicated by a lower error of the first two Betti numbers (β_2 indicating the difference in the number of cavities is always 0 in two dimensions). Moreover, we find that using our skeletonization methods yield slightly better results than using a morphological skeletonization algorithm. Spatial accuracy, quantified by the Dice similarity coefficient (DSC), is nearly identical in all cases.

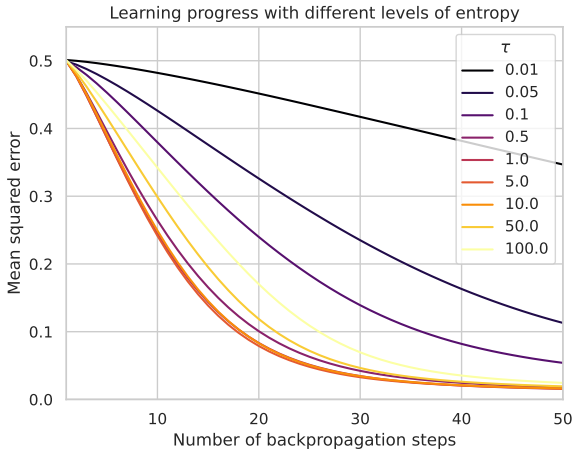

Figure 3. Effect of the Boltzmann temperature τ on the ability to propagate a gradient through our skeletonization module. Both very low entropy and very high entropy inhibit learning.

References

[1] Gilles Bertrand et al. Powerful parallel and symmetric 3d thinning schemes based on critical kernels. *Journal of Mathematical Imaging and Vision*, 48:134–148, 2014. 3

[2] Ilke Demir et al. Skelneton 2019: Dataset and challenge on deep learning for geometric shape understanding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and*

Table 2. Performance of the vessel segmentation network using either a standard Dice loss ('Without') or clDice loss with one of three skeletonization algorithms.

Skeletonization	DSC	β_0 error	β_1 error
Without	0.79±0.01	135.6±5.7	23.3±2.1
Morphological [4]	0.79±0.01	58.2±2.1	21.3±1.2
Ours (Euler)	0.79±0.01	58.3±18.6	18.6±0.5
Ours (Boolean)	0.79±0.01	49.6±8.5	20.6±2.1

Pattern Recognition Workshops, 2019. 2

[3] Diederik P Kingma et al. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 1

[4] Petros Maragos et al. Morphological skeleton representation and coding of binary images. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 34(5):1228–1244, 1986. 4

[5] Nam Hoang Nguyen. U-net based skeletonization and bag of tricks. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2105–2109, 2021. 1, 3

[6] Oleg Panichev et al. U-net based convolutional neural network for skeleton extraction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2019. 1, 3

[7] Olaf Ronneberger et al. 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, 2015. 1

[8] Suprosanna Shit et al. 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, 2021. 1, 3

[9] Joes Staal et al. Ridge-based vessel segmentation in color images of the retina. *IEEE transactions on medical imaging*, 23(4):501–509, 2004. 1

[10] Mihail Ivilinov Todorov et al. Machine learning analysis of whole mouse brain vasculature. *Nature methods*, 17(4):442–449, 2020. 1

[11] Mario Viti et al. Coronary artery centerline tracking with the morphological skeleton loss. In *Proc. ICIP*, pages 2741–2745, 2022. 3