DEEP ELASTICA FOR IMAGE SEGMENTATION

Xu Chen[*1], *Xiangde Luo*[*2], *Guotai Wang*[†2], *Yalin Zheng*[†1]

[1] Department of Eye and Vision Science,
University of Liverpool
[2] School of Mechanical and Electrical Engineering,
University of Electronic Science and Technology of China

ABSTRACT

Image segmentation is a fundamental topic in image processing and has been studied for many decades. Deep learning-based supervised segmentation models have achieved state-of-the-art performance, but most of them are limited by using pixel-wise loss functions for training without geometrical constraints. Inspired by the Euler's Elastica model and recent active contour models introduced into deep learning, we propose a novel active contour with an elastic (ACE) loss function. ACE loss function incorporates Elastica knowledge as geometrically-natural constraints for the image segmentation tasks. In ACE loss function, we introduce the mean curvature, i.e. the average of all principal curvatures, as a more compelling image prior to representing curvature. Furthermore, based on the mean curvature definition, we propose a fast solution (Fast-ACE) to approximate our ACE loss with Laplace operators for three-dimensional (3D) image segmentation. We evaluate our ACE loss and Fast-ACE loss functions on two 2D and 3D biomedical image datasets. Our results show that the proposed loss function outperforms other mainstream loss functions on different segmentation networks. Our source code is available at https://github.com/HiLab-git/ACELoss.

1. INTRODUCTION

Image segmentation is a challenging problem in image processing and has been studied for many decades. The snake/active contour model (ACM) was first proposed by Kass et al. [1] that converts image segmentation problems into energy minimisation problems where the energy of snake/active contours is optimised the object's boundaries. Following that, *Active contour without edge* model a.k.a *Chan-Vese* (CV) model [2] has been widely developed in the past two decades [3, 4] which can be formulated as below,

$$E_{CV}(\phi, c_1, c_2) = \int_{\Omega} |\nabla H_{\epsilon}(\phi)| + \lambda R \tag{1}$$

Where the first term of Eq. 1 is the length of the active contour. The second term is the inside and outside regions of the contour: $R = \int_{\Omega}(c_1 - f)^2 H_{\epsilon}(\phi) + (c_2 - f)^2(1 - H_{\epsilon}(\phi))$. $\Omega \subset R^n$ is a closed subset of the image f to be segmented. c_1, c_2 are the mean value of the inside (foreground) and outside (background) regions respectively. ϕ is a level-set function where the zero level curve represents the segmentation boundary. H_{ϵ} is a smooth approximation of the Heaviside function. λ is a positive hyper-parameter to control the balance between the two terms. In the CV model, the level-set method is often involved in optimising the model by solving partial differential equations (PDEs) iteratively. To tackle the local minimum problem in solving CV models, *fast global minimisation-based active contour model* (FGM-ACM) was proposed by Bresson et al. [5]. FGM-ACM model obtains a global minimum of the ACM with a dual formulation of the total variation (TV) norm.

Recently, Euler's Elastica model is employed in the CV model to segment elongated structures by Zhu et al. [6]. This model will denote as CVE model hereafter for brevity. The CVE model can express as the minimisation of the following functional,

$$E_{CVE}(\phi, c_1, c_2) = \int_{\Omega} \left[\alpha + \beta\Big(\nabla \cdot \frac{\nabla \phi}{|\nabla \phi|}\Big)^2\right] |\nabla H_{\epsilon}(\phi)| + \lambda R \tag{2}$$

Where α and β are positive parameters to control the trade-off between the length and curvature of the segmentation boundary, CVE model has several intrinsic features compared to the CV model:

- Introducing curvature profile
- Preserving connectedness by connecting broken parts of segmentation object to form a meaningful segmentation object
- Reducing missing boundaries interpolation for tubular/curvilinear structures

However, the common challenge for CVE model is computational complexity in numerical schemes because it requires iterative approaches to solve high order PDEs. Inspired by the general idea of CVE and recent deep learning-based ACMs

*Contribution equally
†Corresponding authors

[7, 8], our proposed ACE loss function fuses Elastica as a geometric constraint into a deep learning-based model for image segmentation. Compared to the CVE model, the number of parameters needs to be converged decreased in our model due to supervised learning benefits.

2. METHODOLOGY

Our ACE loss function is defined into discrete form in the case of 3D images as follows,

$$\begin{aligned} E_{ACE}(y,\hat{y}) = (\alpha + \beta\bar{K}^2)\,|\nabla\hat{y}| \\ + \lambda\left|\sum_{i=1}^{H}\sum_{j=1}^{W}\sum_{k=1}^{D}\hat{y}_{i,j,k}(c_1 - y_{i,j,k})^2\right| \\ + \lambda\left|\sum_{i=1}^{H}\sum_{j=1}^{W}\sum_{k=1}^{D}(1-\hat{y}_{i,j,k})(c_2 - y_{i,j,k})^2\right| \end{aligned} \tag{3}$$

The binary ground truth mask and the predicted segmentation denote as y, $\hat{y} : H, W, D \rightarrow R^3$, respectively. $\bar{K}$ is the curvature of $\hat{y}$; c_1 and c_2 are the mean intensity of the inside (foreground) and outside (background) regions respectively and can be defined as constants in advance as $c_1 = 1$ and $c_2 = 0$ [8]; λ is usually set to 1. In our ACE loss, we define $\bar{K}$ by estimating $mean\ curvature$ [9, 10]. That can be a useful and physically-natural constraint to provide a more precise curvature by taking all principal curvatures [11]. Although $mean\ curvature$ has attractive features, minimising $mean\ curvature$-based Elastica regularisation term from the classical CVE model shown in Eq. 2 is still far from practical given enormous computational resources. In the classical CVE model, it leads to a fourth-order PDE [12, 10], even involving other efficient solvers, such as the fixed-point method [13] and the multi-grid method [14], cannot reduce the computational complexity to a satisfactory level for practical purposes. The $mean\ curvature$ can be derived from Monge patch in the case of 3D images as follows,

$$\bar{K} = \frac{\kappa_1 + \kappa_2 + \kappa_3}{3} = \frac{\chi}{\sqrt{1 + \hat{y}_i^2 + \hat{y}_j^2 + \hat{y}_k^2}} \tag{4}$$

Where $\kappa_1, \kappa_2, \kappa_3$ are the three principal curvatures (eigenvalues) at a given point on a surface; $\chi = \hat{y}_{ii}(1 + \hat{y}_j^2 + u_k^2) + \hat{y}_{jj}(1 + \hat{y}_i^2 + \hat{y}_k^2) + \hat{y}_{kk}(1 + \hat{y}_i^2 + \hat{y}_j^2) - 2(\hat{y}_i\hat{y}_j\hat{y}_{ij} + \hat{y}_i\hat{y}_k\hat{y}_{ik} + \hat{y}_j\hat{y}_k\hat{y}_{jk})$; i, j and k are three different directions respectively in the case of 3D images. In the case of 2D images, $mean\ curvature$ can be derived as shown in Eq. 5,

$$\bar{K}_{2D} = \frac{(1 + \hat{y}_i^2)\hat{y}_{jj} + (1 + \hat{y}_j^2)\hat{y}_{ii} - 2\hat{y}_i\hat{y}_j\hat{y}_{ij}}{2(1 + \hat{y}_i^2 + \hat{y}_j^2)^{3/2}} \tag{5}$$

Where $\hat{y}_i$, $\hat{y}_{ii}$, $\hat{y}_{ij}$ and the rest are all approximately computed by central finite differences into discrete form in practice. However, calculating $\bar{K}$ in 3D by Eq. 4 is compute-intensive and time-consuming. The Laplace operator or Laplacian often use for edge detection by calculating the unmixed approximations of the second-order derivatives with discrete and integer-valued convolution kernels to each direction. It is safe to assume that the shape will have small change, then the first-order derivatives will be approximated as small as zero. Laplacian can introduce to simplify the numerator of Eq. 4 by directly computing $\hat{y}_{ii}$, $\hat{y}_{jj}$, $\hat{y}_{kk}$ by kernels to boost DNNs training speed. Our Laplacian-based 3D $curvature$ approximation can express as follows,

$$\bar{K} \approx \nabla^2 = \hat{y}_{ii}^2 + \hat{y}_{jj}^2 + \hat{y}_{kk}^2 \tag{6}$$

Where ∇ is the nabla operation, ∇^2 is the Laplace operator. In practice, Laplacian thus can be estimated by convolution. Our Laplacian-based 3D ACE loss function, denoted as Fast-ACE, has following advantages, 1) it is relatively inexpensive in terms of computations because the tremendous amount of work in 3D convolving operations could be computed by GPUs during the training steps. 2) the 3D *mean curvature* of the shape can be driven efficiently at a minimum of computations loss from the operators.

3. EXPERIMENTS

3.1. Experimental setting

Datasets: In order to evaluate the performance of our ACE loss function, we have applied it to two biomedical image datasets: (1) 2D Digital Retinal Images for Vessel Extraction (DRIVE) retina vessel [15]: DRIVE contains 40 colour fundus images. The 40 resized images (448×448) were divided into 50% of images for training, 25% for validation and the rest of the images for testing. (2) 3D CT image of Pancreas from the Medical Segmentation Decathlon (MSD) [16]: This dataset consists of 281 3D abdominal CT images. We randomly selected 200 cases for the training, 20 for validation and 61 for testing. Following [17], we used the soft tissue CT window range of [-125, 275]. Then, we re-sampled all the cases to the resolution of $1.0 \times 1.0 \times 3.0\ mm^3$. Finally, we cropped all images centring at the pancreas region and normalized them to zero mean and unit variance.

Network architectures: To investigate the robustness and generalizability of the ACE loss function, U-Net [18] and Context Encoder Network (CE-Net) [19] are used as the 2D segmentation networks whilst 3D U-Net, and V-Net [20] as our 3D segmentation networks. All the above networks are not pre-trained on any image datasets.

Training and inference: All the models were implemented by using Python 3.7 and PyTorch 1.4.0. All training experiments were done via one node of a cluster with sixteen 8-core Intel CPUs, 8 TESLA V100 GPUs and 1TB memory. The batch size and total training epochs were set as 8 and 600 respectively. All the models were trained by using Adam

optimizer. For a fair comparison of different loss functions, we searched the optimal learning rate in [e^{-1}, e^{-2}, e^{-3}, e^{-4}] for each loss function respectively based on the validation set [21]. We used the standard on-the-fly data augmentation methods to enlarge dataset and avoid over-fitting during the training stage. In the inference phase, we did not use any post-processing method to boost the performance for a fair comparison.

Evaluation metrics: Two widely used metrics, dice coefficient score (DSC) and the 95^{th} percentile of Hausdorff Distance (HD_{95}) are used to evaluate the segmentation results quantitatively.

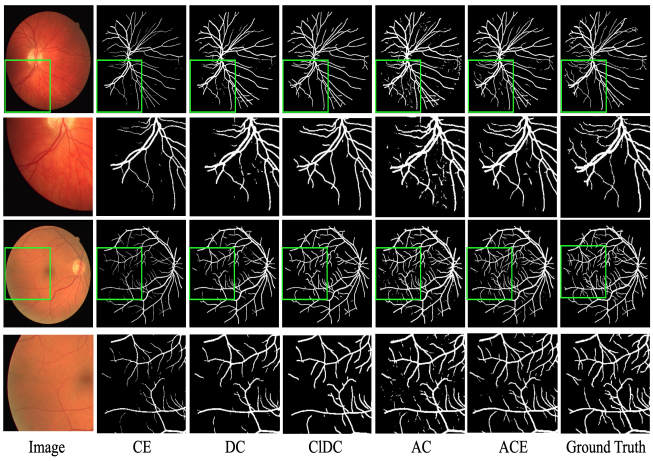

Fig. 1. Example 2D segmentation results of our proposed ACE loss function compared to the other loss functions. The 2nd and 4th row are zoomed retinal vessel segmentation results cropped by the green rectangles from the 1st and 3rd row. From left to right are the original image, CE, DC, ClDC, AC, ACE loss functions and ground truth, respectively.

3.2. Ablation study

We first investigated the optimal value of regularization weights α and β in our proposed ACE loss in Eq. 3 based on all validation datasets with 2D- and 3D- U-Net respectively. We introduced a variable-controlling method to perform this ablation study to investigate the individual impact of different α and β values. Firstly, we fixed the β to 1 to investigate the impact of α for ACE loss function performance. Then, we set the α to 0.001 in all experiments to investigate the impact of β on model performance. The results of the segmentation performance in terms of DSC and HD_{95} on 2D DRIVE validation set and 3D Pancreas CT validation set when α and β are with different values, show that increasing α from 0 to 0.001 leads to improved performance in the 2D α ablation experiment. When α is greater than 0.001, the segmentation performance in 2D decreases gradually. In the 2D β ablation experiment, the performance of β=1 is better than the performance of β=0 (the *curvature* constant is not involved) in terms of HD_{95}. When β is larger than 2, the segmentation performance decreases significantly. The α and β of ACE loss affect the 3D Pancreas segmentation performance similarly. Increasing the α from 0 to 0.1 leads to improved performance, when α is larger than 0.1, the segmentation performance worse rapidly. For the value of β, it can be observed that the performance of β=1 is better than the performance of β=0 (the *curvature* constant is not involved) in terms of DSC. Based on our observation from the above ablation study, our proposed ACE loss function has potential to be deployed in different DNNs-based image segmentation tasks by using well-chosen α in the range of [0.0001, 0.1] and the β values in the structure-wised range of (0, 10]. It has the best performance on 2D DRIVE when α and β are set to 0.001 and 2, respectively. For the 3D Prances data, it has the best performance when α = 0.001 and β = 10. All the α and β were set as optimal values in the following experiments.

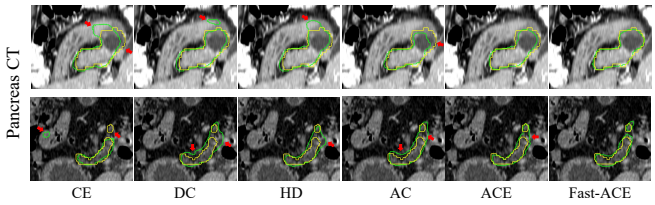

Fig. 2. Example 3D segmentation results of our proposed ACE and Fast-ACE loss function compared to the other loss functions. The first and second row are pancreas segmentation results. The green and yellow contours denote the segmentation and the ground truth, respectively. Red arrows highlight some mis-segmentation

3.3. Comparison to other loss functions

We further compared the performance of the ACE loss with five widely used loss functions when appropriate: CE [18], DC [20], ClDC [22], AC [8]. We evaluated our ACE loss's extendibility and effectiveness by training four different DNNs (U-Net and CE-Net for 2D curvilinear segmentation; 3D U-Net and V-Net for 3D organ segmentation).

Curvilinear structure segmentation from 2D images: We compared ACE loss with CE, DC, ClDC and AC functions. CE loss and DC loss are the most widely used loss functions for image segmentation. ClDC loss and AC loss are proposed recently incorporating morphological skeletons and active contours to enable DNNs focus on the objects' geometric information. ACE loss achieved better performance than all the other loss functions in terms of DSC of 0.833±0.019 (U-Net) and 0.806±0.020 (CE-Net), HD_{95} of 4.068±1.719 pixels (U-Net) and 5.802±2.232 pixels (CE-Net), respectively. Figure 1 shows four segmentation results of using U-Net with five different loss functions, highlighting the ACE loss can

Table 1. Quantitative 2D and 3D segmentation results (mean ± standard deviation) of our proposed ACE and Fast-ACE loss functions compared to other loss functions. CE [18], DC [20], ClDC [22], AC [8] and our ACE and Fast-ACE loss function are evaluated on the DRIVE and Pancreas CT datasets, respectively.

Objects	Model (Network + loss)	DSC ↑	HD_{95} ↓	Time ↓ (s / epoch)	Model (Network + loss)	DSC ↑	HD_{95} ↓	Time ↓ (s / epoch)
Retinal Vessels	U-Net+CE	0.413±0.055	10.350±22.655	≈ 2.3	CE-Net+CE	0.152±0.055	71.875±47.765	≈ 2.5
	U-Net+DC	0.571±0.046	7.107±26.964		CE-Net+DC	0.594±0.042	7.624±26.874	
	U-Net+ClDC	0.587±0.039	6.672±27.046		CE-Net+ClDC	0.675±0.043	6.801±28.298	
	U-Net+AC	0.640±0.039	6.362±27.159		CE-Net+AC	0.664±0.040	6.149±27.947	
	U-Net+ACE	**0.675±0.034**	**5.742±28.425**		CE-Net+ACE	**0.683±0.041**	**5.902±28.281**	
Pancreas	U-Net+CE	0.804±0.068	6.522±3.965	77.2	V-Net+CE	0.796± 0.081	7.239±4.943	76.5
	U-Net+DC	0.813±0.06	6.164±3.674	64.3	V-Net+DC	0.791±0.085	6.703±4.123	61.8
	U-Net+HD	0.816±0.062	6.182±3.749	183.0	V-Net+HD	0.816±0.078	6.164±4.041	176.5
	U-Net+AC	0.828±0.064	6.243±4.773	59.7	V-Net+AC	0.819±0.096	6.151±5.307	55.5
	U-Net+ACE	0.835±0.059	5.521±3.298	78.1	V-Net+ACE	0.827±0.072	6.124±4.494	75.2
	U-Net+Fast-ACE	**0.837±0.059**	**5.481±3.354**	65.5	V-Net+Fast-ACE	**0.832±0.070**	**6.013±5.314**	61.5

preserve curvilinear structure connectedness.

Segmentation of 3D CT images: We compared our proposed ACE, and Fast-ACE loss functions with CE, DC, HD and AC loss functions on the pancreas CT dataset and the results are presented in Table 1. For the pancreas segmentation, the Fast-ACE loss with 3D U-Net and V-Net outperform existing loss functions in terms of DSC of 0.832±0.070 (U-Net) and 0.837±0.059 (V-Net), HD_{95} of 6.013±5.314 voxels (U-Net) and 5.481±3.354 voxels (V-Net) respectively. For the computing time during the training stage, the Fast-ACE loss with U-Net and V-Net spend 65.5s and 61.5s per epoch on the pancreas dataset, shorter than ACE loss with U-Net and V-Net in 78.1s and 75.2s per epoch, respectively. There is almost no difference in predicting the same network as the loss function will not be used. Figure 2 presents some segmentation results of different pancreas dataset methods (all results obtained by V-Net). It can be observed that the segmentation results of the proposed ACE loss and Fast-ACE loss are more accurate compared with the other existing methods.

4. DISCUSSION & CONCLUSION

In this work, we proposed and implemented a new ACE loss for DNNs-based end-to-end image segmentation. Compared to standard ACMs that require an iterative approach to solve PDEs for each image, supervised DNNs will hugely reduce the computational time on segmenting new images after the training. We found that the regularization weight α in the Elastica constraints can be fixed to 0.001 for different image segmentation tasks. The regularization weight β is more sensitive to different images and objects structures. For instance, for curvilinear or tubular structures image segmentation tasks, a β ($0 < \beta < 2$) has a better segmentation result whilst a β ($2 < \beta < 10$) for non-tubular structures. Notably, from the quantitative comparison of segmentation performance between AC and our ACE loss, we observe that our Elastica constraint is more effective to use the geometrical information to constrain segmentation process than the length constraint only of the AC loss, which leads to improved segmentation results in terms of DSC and HD_{95} with four different networks on four different image datasets. For the learning efficiency, when The number of the training set is relatively large, (e.g. the Pancreas dataset), the proposed Fast-ACE loss can maintain similar results and improve training efficiency 20% (this improvement may not be significant for small dataset). The advantage of this new loss function is that it can seamlessly integrate the geometrical information (e.g. curvature and length of the target shape) with region similarity, thus leading to more accurate and reliable segmentation. We introduced mean curvature as a more precise image prior to representing curvature in our ACE loss. Based on the mean curvature definition, we propose a fast 3D solution Fast-ACE to speed up the training process for 3D image segmentation. We applied both the ACE and Fast-ACE to two datasets, and the results showed that they outperform state-of-the-art loss functions.

5. REFERENCES

[1] Michael Kass, Andrew Witkin, and Demetri Terzopoulos, "Snakes: Active contour models," *International Journal of Computer Vision*, vol. 1, no. 4, pp. 321–331, 1988.

[2] Tony Chan and Luminita Vese, "An active contour model without edges," in *International Conference on Scale-Space Theories in Computer Vision*. Springer, 1999, pp. 141–151.

[3] Lei Wang, Guangqiang Chen, Dai Shi, Yan Chang, Sixian Chan, Jiantao Pu, and Xiaodong Yang, "Active contours driven by edge entropy fitting energy for image segmentation," *Signal Processing*, vol. 149, pp. 27–35, 2018.

[4] Tony F Chan, B Yezrielev Sandberg, and Luminita A Vese, "Active contours without edges for vector-valued

images," *Journal of Visual Communication and Image Representation*, vol. 11, no. 2, pp. 130–141, 2000.

[5] Xavier Bresson, Selim Esedoḡlu, Pierre Vandergheynst, Jean-Philippe Thiran, and Stanley Osher, "Fast global minimization of the active contour/snake model," *Journal of Mathematical Imaging and Vision*, vol. 28, no. 2, pp. 151–167, 2007.

[6] Wei Zhu, Xue-Cheng Tai, and Tony Chan, "Image segmentation using euler's elastica as the regularization," *Journal of Scientific Computing*, vol. 57, no. 2, pp. 414–438, 2013.

[7] Shir Gur, Lior Wolf, Lior Golgher, and Pablo Blinder, "Unsupervised microvascular image segmentation using an active contours mimicking neural network," in *Proceedings of the IEEE International Conference on Computer Vision*, 2019, pp. 10722–10731.

[8] Xu Chen, Bryan M Williams, Srinivasa R Vallabhaneni, Gabriela Czanner, Rachel Williams, and Yalin Zheng, "Learning active contour models for medical image segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 11632–11640.

[9] Jean E Taylor, "Ii—mean curvature and weighted mean curvature," *Acta Metallurgica et Materialia*, vol. 40, no. 7, pp. 1475–1485, 1992.

[10] Yuanhao Gong and Orcun Goksel, "Weighted mean curvature," *Signal Processing*, vol. 164, pp. 329–339, 2019.

[11] Yuanhao Gong, *Spectrally regularized surfaces*, Ph.D. thesis, ETH Zurich, 2015.

[12] Wei Zhu and Tony Chan, "Image denoising using mean curvature of image surface," *SIAM Journal on Imaging Sciences*, vol. 5, no. 1, pp. 1–32, 2012.

[13] Fenlin Yang, Ke Chen, Bo Yu, and Donghui Fang, "A relaxed fixed point method for a mean curvature-based denoising model," *Optimization Methods and Software*, vol. 29, no. 2, pp. 274–285, 2014.

[14] Carlos Brito-Loeza and Ke Chen, "Multigrid algorithm for high order denoising," *SIAM Journal on Imaging Sciences*, vol. 3, no. 3, pp. 363–389, 2010.

[15] Joes Staal, Michael D Abràmoff, Meindert Niemeijer, Max A Viergever, and Bram Van Ginneken, "Ridge-based vessel segmentation in color images of the retina," *IEEE Transactions on Medical Imaging*, vol. 23, no. 4, pp. 501–509, 2004.

[16] Amber L Simpson, Michela Antonelli, Spyridon Bakas, Michel Bilello, Keyvan Farahani, Bram Van Ginneken, Annette Kopp-Schneider, Bennett A Landman, Geert Litjens, Bjoern Menze, et al., "A large annotated medical image dataset for the development and evaluation of segmentation algorithms," *arXiv preprint arXiv:1902.09063*, 2019.

[17] Yuyin Zhou, Zhe Li, Song Bai, Chong Wang, Xinlei Chen, Mei Han, Elliot Fishman, and Alan L Yuille, "Prior-aware neural network for partially-supervised multi-organ segmentation," in *Proceedings of the IEEE International Conference on Computer Vision*, 2019, pp. 10672–10681.

[18] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-assisted Intervention*. Springer, 2015, pp. 234–241.

[19] Zaiwang Gu, Jun Cheng, Huazhu Fu, Kang Zhou, Huaying Hao, Yitian Zhao, Tianyang Zhang, Shenghua Gao, and Jiang Liu, "Ce-net: context encoder network for 2d medical image segmentation," *IEEE Transactions on Medical Imaging*, vol. 38, no. 10, pp. 2281–2292, 2019.

[20] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *2016 Fourth International Conference on 3D Vision (3DV)*. IEEE, 2016, pp. 565–571.

[21] Davood Karimi and Septimiu E Salcudean, "Reducing the hausdorff distance in medical image segmentation with convolutional neural networks," *IEEE Transactions on Medical Imaging*, 2019.

[22] Suprosanna Shit, Johannes C Paetzold, Anjany Sekuboyina, Andrey Zhylka, Ivan Ezhov, Alexander Unger, Josien PW Pluim, Giles Tetteh, and Bjoern H Menze, "cldice–a topology-preserving loss function for tubular structure segmentation," *preprint arXiv:2003.07311*, 2020.