Topology-Preserving Segmentation Network: A Deep Learning Segmentation Framework for Connected Component

Han Zhang[1] and Lok Ming Lui[1]

Dept. of Mathematics, The Chinese University of Hong Kong,
Hong Kong, China

Abstract. Medical image segmentation, which aims to automatically extract anatomical or pathological structures, plays a key role in computer-aided diagnosis and disease analysis. Despite the problem has been widely studied, existing methods are prone to topological errors. In medical imaging, the topology of the structure, such as the kidney or lung, is usually known. Preserving the topology of the structure in the segmentation process is of utmost importance for accurate image analysis. In this work, a novel learning-based segmentation model is proposed. A *topology-preserving segmentation network (TPSN)* is trained to give an accurate segmentation result of an input image that preserves the prescribed topology. TPSN is a deformation-based model that yields a deformation map through a UNet, which takes the medical image and a template mask as inputs. The main idea is to deform a template mask describing the prescribed topology by a diffeomorphism to segment the object in the image. The topology of the shape in the template mask is well preserved under the diffeomorphic map. The diffeomorphic property of the map is controlled by introducing a regularization term related to the Jacobian in the loss function. As such, a topology-preserving segmentation result can be guaranteed. Furthermore, a multi-scale TPSN is developed in this paper that incorporates multi-level information of images to produce more precise segmentation results. To evaluate our method, we applied the 2D TPSN on Ham10000 and 3D TPSN on KiTS21. Experimental results illustrate our method outperforms the baseline UNet segmentation model with/without connected-component analysis (CCA) by both the dice score and IoU score. Besides, results show that our method can produce reliable results even in challenging cases, where pixel-wise segmentation models by UNet and CCA fail to obtain accurate results.

1 Introduction

Automated anatomical segmentation is an important procedure for many clinical applications, such as diagnosing diseases and monitoring disease progression. Recent developments in deep learning have achieved great success in medical image segmentation. Once successfully trained, the deep neural network can segment the input image in real-time. Nevertheless, the major challenge is the

accuracy of the segmentation result. Accurate segmentation is essential for meaningful medical image analysis. In particular, the accurate topology of the segmented anatomical structure is crucial for the measurement of the structure. Unfortunately, most existing algorithms are prone to topological errors, causing mis-segmentation with multiple components or thin connections. In practical situations, the topologies of anatomical structures are often known. For instance, structures like the kidney or liver are known to be simply-connected. We are thus motivated to develop a deep segmentation model with a topological prior that learns to segment with correct topology. State-of-the-art deep segmentation approaches commonly assign a label to each pixel and can be formulated as a pixel-wise classification problem. Under this formulation, the topological constraint is hard to be enforced. Deformation-based models[1,12,17,16], which deform a template mask by a suitable deformation map to segment the image, have been recently proposed. The topological prior can be easily incorporated using this approach, which will be adopted in this work.

In this paper, a learning-based topology-preserving segmentation framework that learns to segment an input image with a given topological prior is proposed. TPSN takes the image to be segmented and a template mask describing the prescribed topology as inputs. It outputs a diffeomorphic deformation map that deforms the template mask to segment the input image. For instance, suppose the kidney in an image is to be extracted, the template shape can be chosen as a simple disk, and will then be deformed by the diffeomorphism to enclose the kidney in the input image. The diffeomorphic deformation map ensures the topology of the segmented shape is consistent with that of the template. The diffeomorphic property of the map is enforced by regularizing the Jacobian and the Laplacian of the map. With the trained network, the segmentation process can be done efficiently and the topology of the segmented shape is guaranteed to be consistent with the prescribed topological prior. To further improve the segmentation result, a *multi-level topology-preserving segmentation network (mlTPSN)*, which incorporates multi-level information by producing masks from coarse to fine, is introduced. At the low-fidelity level, a low-resolution version of the image is fitted into the network to predict a rough mask. In the higher levels, rough masks are used as the template mask to generate a more precise mask to segment the higher-resolution image. This multi-level strategy is proved to be effective to enhance the accuracy of the segmentation result. To evaluate the proposed TPSN and mlTPSN, experiments are carried out on both 2D and 3D medical images. Our methods outperform the baseline UNet with and without connected component analysis. In challenging situations when the UNet segmentation fails, our proposed networks can still yield satisfying and reliable results. These demonstrate the efficacy of our proposed models.

Our contribution of this paper is three-fold: (1) We introduced a deep-learning segmentation network that learns to segment an image accurately with correct topology according to the prescribed topological prior. (2) We designed a special regularization, namely the *ϵ-ReLU Jacobian loss*, which can effectively enforce the bijectivity of the deformation map and prevent heavy shrinkage. (3) We proposed

a multi-scale topology-preserving segmentation network that incorporates multi-level information of images to produce more precise results.

2   Related Work

Medical image segmentation has been widely studied. Numerous deep learning based models have been recently proposed. UNet and its variants achieve great success[10,9,3]. In particular, 3D-UNet [3] and VNet [9] proposed by Çiçek *et al.* and Milletari*et al.* respectively are successful for 3D medical image segmentation. In the challenge of KiTS19[4], nnUNet [6] has achieved outstanding results using UNet and a series of practical pre- and post- processing strategies. Besides, [7] proposed to use one UNet to predict an initial mask and apply an addtional one to refine the segmentation result.

Mathematical models for image segmentation have also been extesnively studied. [8] proposed an active contour model that delineates the object boundary. Chan and Vese[2] generalized active contour model using a level set formulation. Shape prior segmentation have been investigated. Segmentation with topological prior has shown to be effective in enhanching the accurary [4]. Chan *et al.* [1] introduced a deformation based segmenation model using quasiconformal maps for topology preserving segmentation. Siu *et al.* [12] introduced to incorporate the dihedral angle in the model for segmentation with partial convexity prior and topological prior. Zhang *et al.* [16] proposed a deformation based topology-preserving segmentation model through registration using the hyperelastic regularization. The model has been further extended to incorporate the convexity prior[17]. Zhou *et al.* [18] propsoed a model, which accounts for the relative location and size prior statistically. Hu *et al.* [5] designed a continuous-valued loss function to enforce the topological constraint. Shit *et al.* [11] introduced clDice to enhance the connectedness in the segmentation for tubular structures. Wyburd *et al.* [15] enforce topology-preserving from given prior by compositing a series of diffeomorphic layers into network architecture.

3   Method

In this section, we describe our proposed learning-based segmentation framework in detail. Conventional segmentation approaches are commonly formulated as a pixel-wise classification problem, with which the topological constraint is hard to be integrated. In this work, our strategy is to train a deep neural network to transform a template mask capturing the prescribed topological information by a diffeomorphism to segment the input image.

Network Architecture The pipeline and architecture of our proposed model are illustrated in Figure 1. The top diagram shows our proposed *Topology-preserving Segmentation Network (TPSN)*. TPSN takes the image I to be segmented and the template masks M_{temp} as the inputs. The template mask is a binary image describing the prior topological information. For instance, if the structure to be
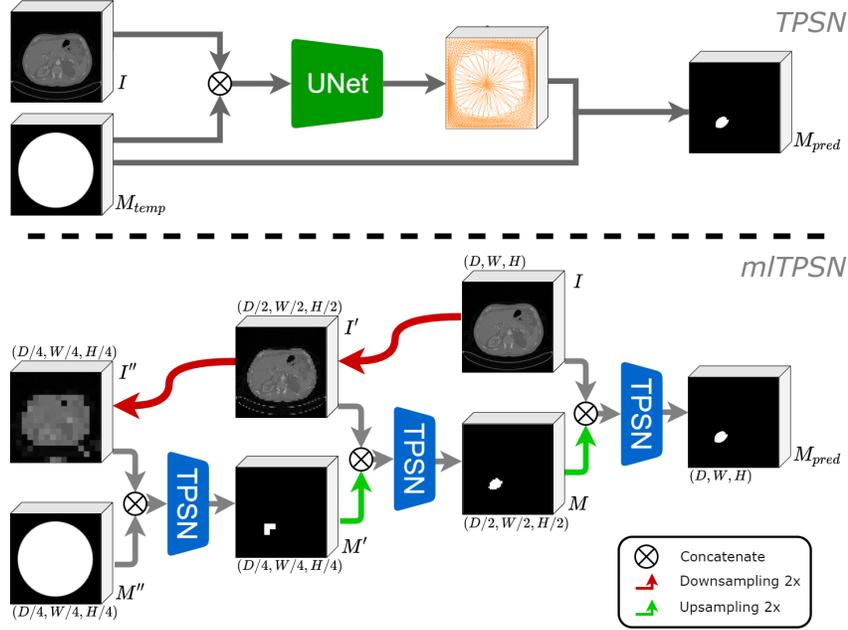
Fig. 1: The framework and architecture for TPSN and multi-level TPSN

extracted is simply-connected, M_{temp} can be chosen as a binary image containing a 2D disk or 3D ball for 2D and 3D segmentation respectively. The network architecture is a UNet. It outputs the deformation mapping f. This output can be regarded as a 2-channel or 3-channel image capturing the coordinates of the deformation mapping for 2D or 3D segmentation respectively. M_{temp} is then deformed by f to obtain a transformed mask M_{pred}, which segments the structure in the image I.

The Loss function To train the network, a suitable loss function is necessary. The loss function $\mathcal{L}$ aims to enforce the diffeomorphic property of the output deformation map, as well as ensuring that M_{pred} is close to the ground truth segmentation result from the training data. $\mathcal{L}$ consists of the fidelity term and regularization term. The fidelity term is chosen as the Dice loss that drives M_{pred} to the ground truth. To enforce the diffeomorphic property, regularization of the Jacobian is performed. A map is diffeomorphic if its Jacobian is positive everywhere [16]. The following ϵ-RELU Jacobian regularizer is applied:

$$\mathcal{L}_{Jac}(f) = ||ReLU_\epsilon(-\det \nabla(f))||_1 \tag{1}$$

where $ReLU_\epsilon(x) = Relu(x + \epsilon)$. $\mathcal{L}_{Jac}$ promotes a positive Jacobian determinant of f that is greater than ϵ everywhere. This prevents flips (non-bijectivity) and heavy shrinkage of f. When ϵ is small, larger deformation can be acheived but more squeezing can be observed. When fine details are to be extracted, ϵ can be tuned to a smaller value. To further enhance the smoothness of f, the Laplacian

loss is adopted:

$$\mathcal{L}_{Lap}(f) = ||\Delta(f)||_1 \tag{2}$$

Then, the final overall loss is as:

$$\mathcal{L} = \lambda_{Dice}\mathcal{L}_{Jac}(f)_{Dice}(M_{output}, M_{label}) + \lambda_{Jac.}\mathcal{L}_{Jac}(f) + \lambda_{Lap.}\mathcal{L}_{Lap}(f) \tag{3}$$

where $\mathcal{L}_{Dice}$,det $\nabla(f)$ and $\Delta(f)$ are Dice loss, the Jacobian determinant of f and Laplacian of f weighted by λ_{Dice}, $\lambda_{Jac.}$ and $\lambda_{Lap.}$ respectively. The derivatives in the Jacobian and Laplacian are computed by forwarding difference and central difference schemes respectively.

Multi-level TPSN To enhance the robustness and accuracy of our proposed model, a multi-level strategy is adopted to build the *multi-level topology-preserving segmentation network (ml-TPSN)*. The pipeline and architecture of ml-TPSN are shown in Figure 1 bottom. The original image I is downsampled into $I^{'}$ and $I^{''}$, which are respectively $1/2$ and $1/4$ of the original dimension. $I^{''}$ is fitted into TPSN with a coarse template mask of the same dimension as $I^{''}$ to output a predicted segmentation mask. The predicted segmentation mask gives a rough approximation of the segmentation. This coarse predicted mask is upsampled and used as the template mask $M^{'}$ for $I^{'}$. $I^{'}$ and $M^{'}$ are fitted into another TPSN to output the predicted segmentation mask for $I^{'}$. The predicted mask is upsampled to M, which is fitted into a TPSN with I to obtain the final segmentation result.

Such a multi-level strategy allows our model to gradually extract features from the image from coarse to fine. A predicted mask is used as the template mask in the next TPSN layer, which is closer to the actual mask. As a result, the algorithm is more robust, especially for structures with complex geometry that are to be segmented. Experiment results show that this multi-level strategy helps to improve the segmentation accuracy with higher segmentation scores.

4 Experiment

4.1 Implementation Details

The implementation details are described in this section. We compare our methods with the baseline UNet, as well as the UNet model with connected component analysis (CCA)[4][6] For UNet with CCA, the simply-connected component with the largest area is chosen as the segmentation result to obtain a simply-connected segmentation result. The UNet model with connected component analysis is denoted by UNet(cca). The experiments of TEDSNet are carried out using the released code of [15].

Dataset We evaluate TPSN and mlTPSN on two datasets, namely, Ham10000[14,13] and KiTS21[4]. Most of the masks provided in Ham10000 are simply connected, which is suitable to evaluate the capability of our method for 2D segmentation. Among 10015 pairs of images and masks in Ham10000, we divide them into 9000 pairs for training and 1015 pairs for testing. All images are downsampled to
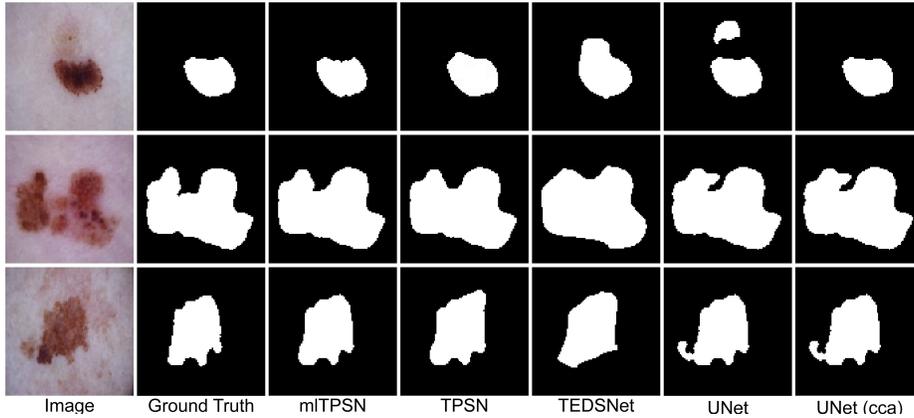
| Image | Ground Truth | mlTPSN | TPSN | TEDSNet | UNet | UNet (cca) |

Fig. 2: Segmentation results comparison for Ham10000

128×128 in both the training and testing process. KiTS21 provides segmentation masks for two instances of Kidney annotation, where the connectedness of individuals is strictly preserved. Restricting one continuous mask region is a reliable and practical prior and post-processing strategy. Amongst 300 cases, we employ 210 cases for training and the other 70 cases for testing. All the masks and volume images are normalized into the same spacing and downsampled into a size of $64 \times 128 \times 128$ after the center is cropped into the same depth. The intensity of images is normalized to $[0, 1]$.

Resources and Parameters All models are trained for 300 epochs with a learning rate of 0.00001 using RMSprop optimizer. The model for 2-dimensional image segmentation is trained with a minibatch of 64 on a CentOS 7 central cluster computing node with one 64GB, 2.4GHz Intel Xeon E5-2680 CPU, and one GeForce GTX 1080 Ti GPU. 3-dimensional volume image segmentation models are trained with a minibatch of 8 images on a node with one 2.4GHz Intel Xeon E5-2680 CPU, and eight GeForce GTX 1080 Ti GPU. The weighting parameters λ_{Dice}, λ_{dice} and λ_{jac} are set to be 1.0,1.0 and 0.1 respectively.

4.2   Experiments on 2D Segmentation

Method	Dice	Best
UNet	93.50 ± 0.29	93.79
UNet(cca)	93.63 ± 0.29	93.92
TEDSNet	89.91 ± 0.27	90.18
TPSN	93.77 ± 0.37	94.14
mlTPSN	94.42 ± 0.39	94.81

Table 1: Result on Ham10000

Method	Dice	Best
UNet	92.76 ± 0.38	93.14
UNet(cca)	92.81 ± 0.38	93.19
TEDSNet	87.67 ± 0.29	88.06
TPSN	92.73 ± 0.48	93.21
mlTPSN	93.17 ± 0.41	93.58

Table 2: Result on KiTS21

We conduct experiments on the Ham10000 dataset to validate the capability of TPSN on labeling continuous masks for 2D images. The results are reported in Table.1. Compared to UNet segmentation model based on a pixel-wise classification, our methods achieve a better result with 94.14% best Dice. Moreover,

mlTPSN achieve a significantly better score with 1.02% higher by Dice (93.79% to 94.81%) than UNet. Using connected component analysis to correct the topological error, the scores are improved to 93.92% by Dice score for the baseline UNet. The results produced by TEDSNet can give a correct topological structure. However, they are not precise and obtain only 90.18% best Dice score.

Figure 2 shows the qualitative visualization. Both TPSN and mlTPSN can avoid topological errors in the first row. As for the results in the second and third rows, the boundaries of the masks predicted by mlTPSN are evidently closer to the ground truth. It demonstrates the effectiveness of the multi-level strategy.

4.3   Experiments on 3D Segmentation

We also test the TPSN and ml-TPSN on KiTS21 dataset to segment the 3D kidney. The experimental results are reported in this subsection.
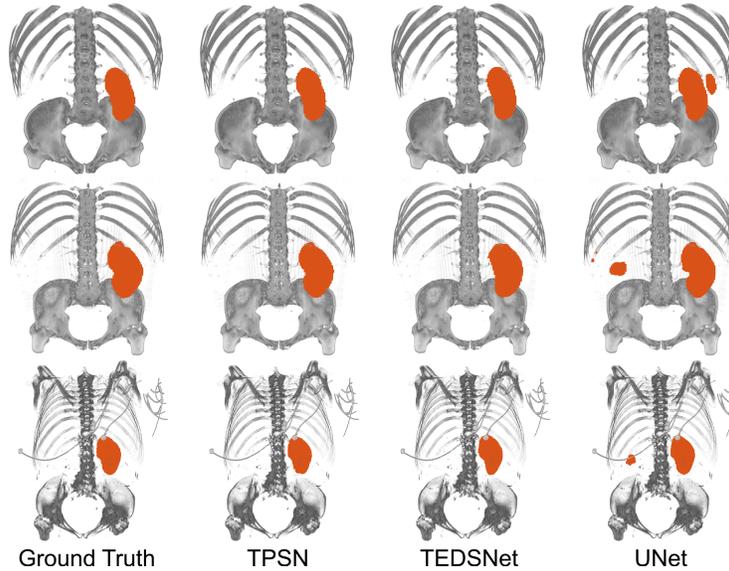

Fig. 3: 3D Segmentation comparison for KiTS21

Figure 3 shows the qualitative visualization results. The masks predicted by our methods are free of outliers and topological errors. On the contrary, the UNet segmentation model mislabels the right kidney while the task is to segment the left one. The quantitative results are reported in Table 2. TPSN achieves a higher score by Dice score, which is 93.21%. Furthermore, our mlTPSN achieves the best result with 93.62% by Dice. It outperforms all other approaches, including the baseline UNet, UNet with connected component analysis, and TEDSNet.
Dataset with Information Loss We also test our models on images with corruption and information loss. This may be brought by patients' movements
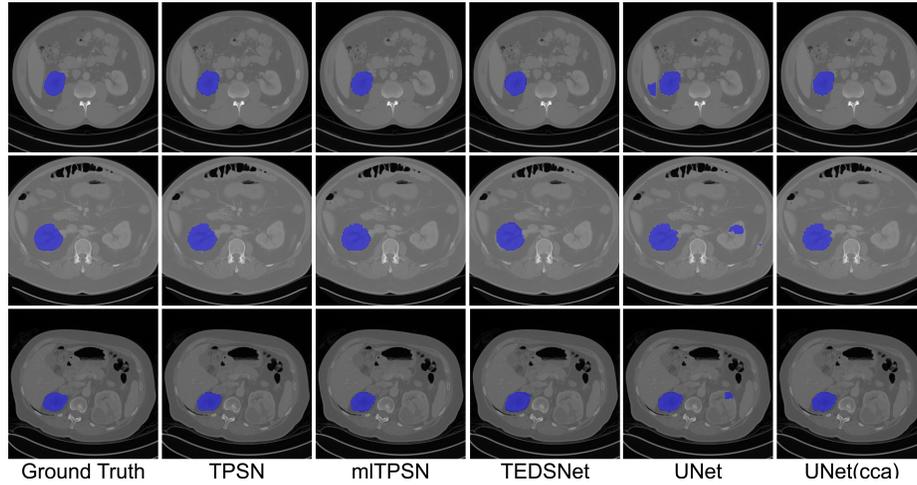
Ground Truth   TPSN   mlTPSN   TEDSNet   UNet   UNet(cca)

Fig. 4: Segmentation results comparison for KiTS21

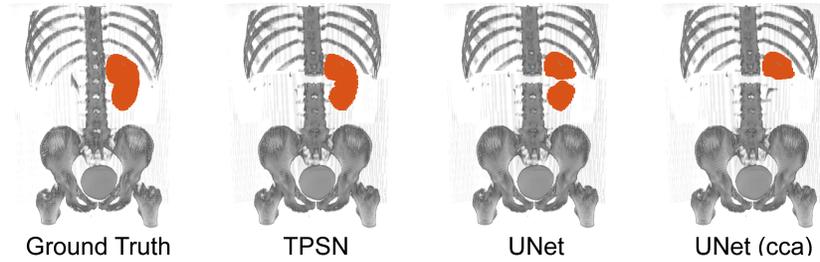Ground Truth   TPSN   UNet   UNet (cca)

Fig. 5: $3D$ Segmentation results comparison for special cases

or device faults of the scanner. In this experiment, we assume the intensity information has been lost for some slices in the 3D image. It is simulated by setting the intensity value of some slices in the volume image of a subject to be zero. The corrupted image is then fitted into the trained network without any further processing. The result is illustrated in Figure 5. It is observed that the result obtained by TPSN is free of topological errors, even though the mask is shrunk a bit compared to the ground truth. On the contrary, both UNet and UNet(cca) fail to give accurate results. It demonstrates the robustness of our proposed models.

5 Conclusion

In this work, we proposed a novel learning-based segmentation framework for both 2D and 3D images that guarantees to preserve the prescribed topology. The main idea is to train a UNet that predicts a diffeomorphic mapping, which registers a template mask to the ground truth. Experimental results show that our methods outperform other methods. Our framework can also be easily incorporated with other architecture to obtain better results.

References

1. Chan, H.L., Yan, S., Lui, L.M., Tai, X.C.: Topology-preserving image segmentation by beltrami representation of shapes. Journal of Mathematical Imaging and Vision **60**(3), 401–421 (2018)
2. Chan, T.F., Vese, L.A.: Active contours without edges. IEEE Transactions on image processing **10**(2), 266–277 (2001)
3. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3d u-net: Learning dense volumetric segmentation from sparse annotation. In: Proc. of Int. Conf. on Medical Image Computing and Computer-Assisted Intervention. pp. 424–432. Springer (2016)
4. Heller, N., Isensee, F., Maier-Hein, K.H., Hou, X., Xie, C., Li, F., Nan, Y., Mu, G., Lin, Z., Han, M., et al.: The state of the art in kidney and kidney tumor segmentation in contrast-enhanced ct imaging: Results of the kits19 challenge. Medical Image Analysis **67**, 101821 (2021)
5. Hu, X., Li, F., Samaras, D., Chen, C.: Topology-preserving deep image segmentation. Proc. of Int. Conf. on Neural Information Processing Systems (2019)
6. Isensee, F., Jaeger, P.F., Kohl, S.A., Petersen, J., Maier-Hein, K.H.: nnu-net: A self-configuring method for deep learning-based biomedical image segmentation. Nature Methods **18**(2), 203–211 (2021)
7. Jha, D., Riegler, M.A., Johansen, D., Halvorsen, P., Johansen, H.D.: Doubleu-net: A deep convolutional neural network for medical image segmentation. In: 2020 IEEE 33rd International symposium on computer-based medical systems (CBMS). pp. 558–564. IEEE (2020)
8. Kass, M., Witkin, A., Terzopoulos, D.: Snakes: Active contour models. International journal of computer vision **1**(4), 321–331 (1988)
9. Milletari, F., Navab, N., Ahmadi, S.A.: V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: Proc. of Int. Conf. on 3D Vision (3DV). pp. 565–571. IEEE (2016)
10. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Proc. of Int. Conf. on Medical Image Computing and Computer-Assisted Intervention. pp. 234–241. Springer (2015)
11. Shit, S., Paetzold, J.C., Sekuboyina, A., Ezhov, I., Unger, A., Zhylka, A., Pluim, J.P., Bauer, U., Menze, B.H.: cldice-a novel topology-preserving loss function for tubular structure segmentation. In: Proc. of IEEE Conf. on Computer Vision & Pattern Recognition. pp. 16560–16569 (2021)
12. Siu, C.Y., Chan, H.L., Ming Lui, R.L.: Image segmentation with partial convexity shape prior using discrete conformality structures. SIAM Journal on Imaging Sciences **13**(4), 2105–2139 (2020)
13. Tschandl, P., Rinner, C., Apalla, Z., Argenziano, G., Codella, N., Halpern, A., Janda, M., Lallas, A., Longo, C., Malvehy, J., et al.: Human-computer collaboration for skin cancer recognition. Nature Medicine **26**(8), 1229–1234 (2020)
14. Tschandl, P., Rosendahl, C., Kittler, H.: The ham10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Scientific Data **5**(1), 1–9 (2018)
15. Wyburd, M.K., Dinsdale, N.K., Namburete, A.I., Jenkinson, M.: Teds-net: Enforcing diffeomorphisms in spatial transformers to guarantee topology preservation in segmentations. In: Proc. of Int. Conf. on Medical Image Computing and Computer-Assisted Intervention. pp. 250–260. Springer (2021)

16. Zhang, D., Lui, L.M.: Topology-preserving 3d image segmentation based on hyperelastic regularization. Journal of Scientific Computing **87**(3), 1–33 (2021)
17. Zhang, D., Tai, X.c., Lui, L.M.: Topology-and convexity-preserving image segmentation based on image registration. Applied Mathematical Modelling **100**, 218–239 (2021)
18. Zhou, Y., Li, Z., Bai, S., Wang, C., Chen, X., Han, M., Fishman, E., Yuille, A.L.: Prior-aware neural network for partially-supervised multi-organ segmentation. In: Proc. of Int. Conf. on Computer Vision. pp. 10672–10681 (2019)