Topology-Preserving Segmentation Network: A Deep Learning Segmentation Framework with Topology Constraint

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

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

Abstract. [1] Do pixel-wise classification on the input image for segmentation task is the conventional way. Through the local and global information can take into the computation through convolution, it can not guarantee the final output masks with any topological property. For objects that are known to be a single connected component, a natural idea is to use a closed scope (for 2D) or an elastic ball (for 3D) to enclose the objects. In this work, we propose a new framework that enables such prior on the topology of the objective to be enforced in deep learning models. Through such prior knowledge, there will be no outliers in the predicted output. Besides using such framework independently, we can also use it to do post-processing for masks by other methods and fuse it into any existing segmentation methods.

1 Introduction

Medical images acquisition follows a strict rule and presents the same type of anatomic organs. Thus, abundant prior knowledge can be provided for medical image segmentation. For example, the relative position and the shape of the objects to be segmented may be similar to each other. Most human organs are known to be single connected component[12]. However, most of the previous work by deep learning accomplish the segmentation task by doing pixel-wise classification. Though with the help of convolution and deep layers of the network, the global and local information can be encoded and processed for segmentation, there is no guarantee that the prior topology property of the objects can be embedded.

In this paper, we introduce a topology-preserving segmentation network that can enforce a strong topology prior knowledge in segmenting process. Specifically, in tasks for segmenting objects that are known to lie in a single-connected component region with no holes within the region, the segmentation masks produced by our method can promise to be free of inside holes and outliers. We do this by using a UNet to predict a transformation mapping for a template mask, which should register the masked approach to the ground truth mask as much

[1] This work is still in progress

as possible. Since the template mask is a single connected component region without holes, the transformed mask should also be after the topology-preserving mapping. In our model, the mapping predicted by the UNet will be enforced to be bijective, that will not change the topology of transformed masks, by carefully designed regularition.From the quantitative results, our method achieves a higher score. By the visualization, results produced by our methods are free of outliers and inner holes.

The contribution of this works are three folds: (1) We introduced a deep learning framework that can guarantee the final output are single-connected components that are free of outliers and inside holes. (2) Our registration-based segmentation method can be fused with any other segmentation method to achieve better results. (3) The masks produced by our methods can be used to do post-processing for masks predicted by other methods.

2 Related Work

Medical image segmentation is the corner brick for applying artificial intelligence to medicine. Neural network model U-net and its variants achieve great success in this field. U-net[6,5,1] its variants. For lots of medical images are three-dimensional volume images, the information should be considered in 3-dimensional space sense. Then, 3D-UNet and VNet are proposed by Çiçek*et al.* and Milletari*et al.* respectively. In the challenge of KiTS19[2], nnUNet [4] achieved significant results using UNet and a series of practical pre- and post- processing strategies.

In the KiTS19 challenge, lots of works use connected-component analysis to do post-processing to remove outliers and inner holes, which is a strong tool brought by the prior on the topology property of organs to be segmented[2]. Such a common choice promises that prior knowledge on topology is a strong tool to promise the results with higher accuracy and precision. Zhang *et al.* [11] proposed a topology-preserving segmentation model through registration and guaranteed the final output to be single-connected. Zhou *et al.* [12] account for the relative location and size prior statistically. Hu *et al.* [3] designed a continuous-valued loss function to enforce the same topology as the ground truth masks. Shit *et al.* [7] introduced clDice to enhance the connectedness in the segmentation for tubular structures.

3 Method

In this section, we will introduce the two main factors in our method. The framework architecture and the carefully designed regularization loss. After that, inspired by the common postprocessing tool, connected component analysis, the Hybrid topology-preserving segmentation network that fuses the advantages of U-Net and TPSN are presented.

The conventional segmentation method is to do pixel-wise classification on the input image, through the local and global information can took into the computation through convolution, it can not guarantee the final output masks

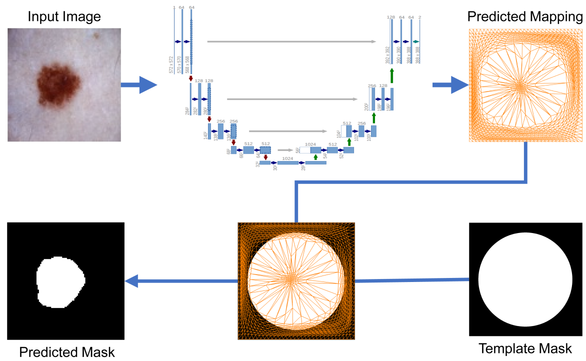

Fig. 1. Caption

with any topological property. In some cases, when the objects are known to be some connected components, a natural idea is to use a closed scope (for 2D) or an elastic ball (for 3D) to enclose the objects. Inspired by this, an interesting approach to do this is to transform a template mask image to register to the ground truth mask by the information provided by the input image.

Module Architecture The pipeline and architecture of this method are visualized in Figure.1. Like other UNet architecture, the images to be segmented are input into the network to predict twothree channel images. The twothree channel images, which can also be referred to as the offset vector field alternatively, will be the transformation mapping for registering the template mask to the ground truth mask. The enclosed region in the transformed template mask is the final segmentation in the task.

The Loss and Jacobian&Laplacian Regularition The objective of the training is to enforce the network to output a mapping f that can transform the template mask to approach the ground truth as much as possible. To achieve this, we make the first term in our loss function the Dice loss. Then, directed by the truth that a smooth mapping is bijective given the Jacobian is positive everywhere[11], we have our second and third regularity by the Jacobian and Laplacian of the mapping. Thus, the final loss function is as:

$$E = \lambda_{Dice} E_{Dice}(M_{output}, M_{label}) + \lambda_{Jac.} \det \nabla(f) + \lambda_{Lap.} \Delta(f) \tag{1}$$

where E_{Dice},$\det \nabla(f)$ and $\Delta(f)$ are Dice loss, the determinant of Jacobian matrix of f and Laplacian of f weighted by λ_{Dice}, $\lambda_{Jac.}$ and $\lambda_{Lap.}$, respectively. The implementation for the direvitives in Jacobian is by forward difference method while it's central difference method for Laplacian.

Hybrid Segmentation Network Point-wisely multiplies the U-Net predicted mask and the TPSN predicted mask.

4 Experiment

4.1 Implementation Details

Dataset We evaluate the proposed topology-preserving segmentation network on two datasets, Skin Cancer MNIST[9,8] and KiTS21[10]. Most of the masks provided in Skin Cancer MNIST are simply connected, which is suitable to evaluate the capability of our method for 2D imaging. Among 10015 pairs of images and masks in Skin Cancer MNIST, we divide them into 9000 pairs for training and 1015 pairs for testing. All the images are downsampled to 128×128 in both the training and testing process. KiTS21 provides segmentation masks for two instances of Kidney annotation mostly, where the connectedness of individuals is strictly kept. Restricting one continuous mask region is a strong and practical prior and post-processing strategy. Among 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 the size of $64 \times 128 \times 128$ after the center is cropped into the same depth. The intensity of images is normalized into $[0, 1]$.
Computing Resources All models are trained for 300 epochs with a learning rate of 0.00001 using RMSprop. 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 parameter λ_{dice} and λ_{jac} is set to be 1.0 and 0.1 respectively.

Method	DSC	mIoU
UNet	93.76	89.18
TPSN	94.07	89.69
Hybrid-UNet	94.89	90.97

Table 1. Result on the skin leision segmentation of Ham10000

Method	DSC	mIoU
UNet	93.06	87.83
TPSN	93.27	89.01
Hybrid-UNet	94.15	90.21

Table 2. Result on the kidney segmentation of KiTS21 (no data augmentation)

4.2 Experiments on 2D Segmentation

We conduct experiments on the Skin Cancer MNIST dataset to validate the capability of TPSN on labeling continuous masks for 2D images. The results are given in Table.4.1. Compared to simply taking U-Net to do segmentation by pixel-wise classification, our methods achieve a better result with 94.07% DSC, and 89.69% IoU. Combining the predicted masks by U-Net and masks by TPSN, we achieve a significantly better score with 1.13% higher by DSC

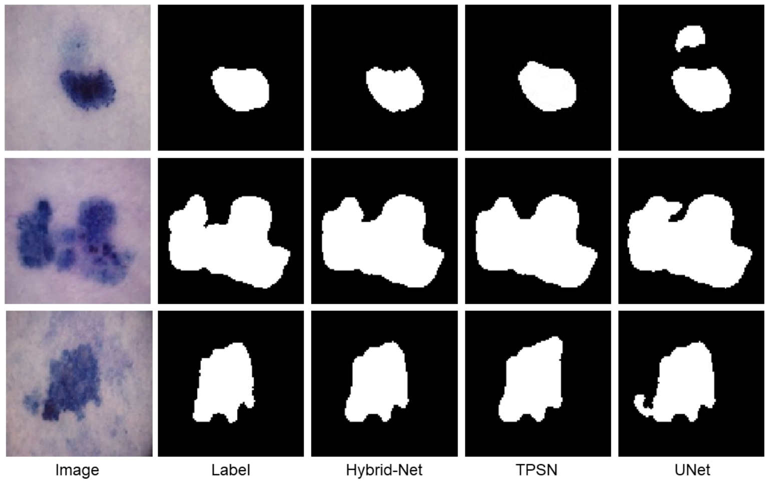

Fig. 2. Caption

(93.76% to 94.89%) and 1.79% in IoU (89.18% to 90.97%). From the qualitative visualization illustrated in Figure.2, the outliers and the isolated masks in the incorrect regions are avoided. In addition, the Hybrid-TPSN keeps the advantages of topology-preserving segmentation by TPSN and the flexibility of pixel-wise segmentation by U-Net, which enables the predicted masks free of outliers and precise in the boundary regions.

4.3 Experiments on 3D Segmentation

Many human organs like kidneys and liver are located in a relatively similar location and with close size. Thus, it's natural and wise to employ these prior pieces of knowledge in the segmentation process as well as data preprocessing and postprocessing. However, a connected organ in 3D may pretend to be two separate parts in axial slices due to its non-convexity. Following these principles, we evaluate our method by segmenting the left kidney from 3D volume images provided by KiTS21.

From the visualization presented in Figure.3, the masks predicted by our methods are free of outliers and enforce the masked regions in a strictly connected manner. Moreover, people who didn't encounter radical nephrectomy have two kidneys distributed on the two sides of the body. The UNet will mislabel the right kidney when the task is to segment the left one. From the visualized results by our method, such a problem will be solved out credibly. Benefitting from these, TPSN achieves a higher score by both DSC and IoU, which is 93.27% and 89.01%, respectively. Besides that, our Hybrid-TPSN can erase the outliers of the masks predicted by other methods like a simple U-Net, and achieve a DSC score of 94.15% and 90.21% IoU, which is 1.09% and 2.38% higher than the results by baseline U-Net.

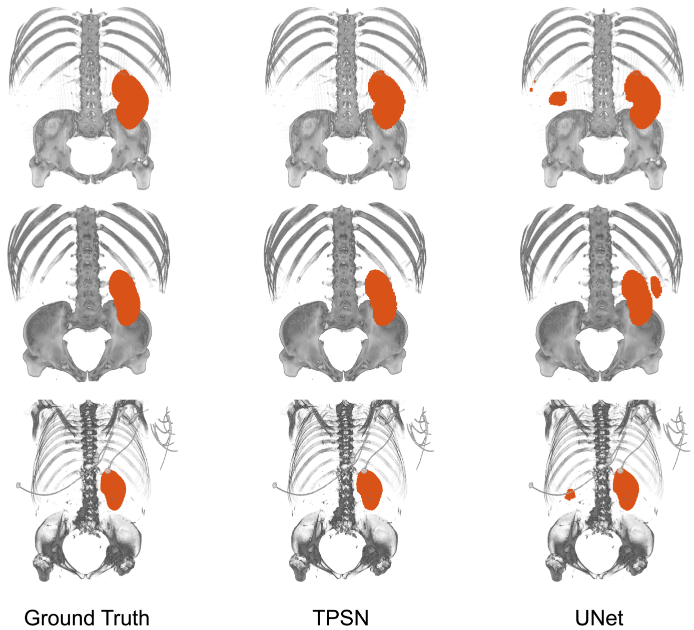

Fig. 3. Caption

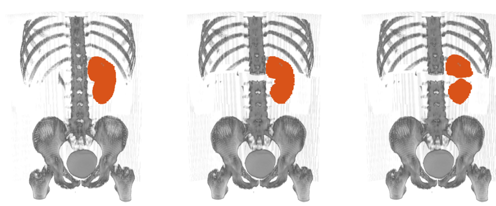

Fig. 4. Caption

A Special Case on Information Loss We also evaluate the capability of our method on the tolerance of noise or information loss, which may be brought by the patient's movement or the device fault of the scanners. Here we assume a 3D volume image lost the intensity information for some slices that contain the regions to be segmented. We simulate this by set the intensity value of some slices in the normalized volume image of a subject to be zero. Then input the processed image into the network we trained in the above without any other processing. From the result illustrated in Figure.4, we can see that even the masks output by TPSN shrink a little bit compare to the ground truth, it's still a single-connected region. However, the masks predicted by UNet are in two seperate parts.

5 Conclusion

In this work, we proposed a new deep learning segmentation framework for both 2D and 3D images that can guarantee to produce single-connected masks without holes. The idea is to train the UNet to predict a mapping that can register a template mask to the ground truth mask. The quantitive and visual results show that our methods achieve better segmentation results. Besides, our model can be easily fused with other methods to produce better results.

References

1. 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)
2. 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)
3. Hu, X., Li, F., Samaras, D., Chen, C.: Topology-preserving deep image segmentation. Proc. of Int. Conf. on Neural Information Processing Systems (2019)
4. 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)
5. 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)
6. 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)
7. 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)
8. 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)
9. 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)
10. Wolny, A., Cerrone, L., Vijayan, A., Tofanelli, R., Barro, A.V., Louveaux, M., Wenzl, C., Strauss, S., Wilson-Sánchez, D., Lymbouridou, R., et al.: Accurate and versatile 3d segmentation of plant tissues at cellular resolution. Elife **9**, e57613 (2020)
11. Zhang, D., Lui, L.M.: Topology-preserving 3d image segmentation based on hyperelastic regularization. Journal of Scientific Computing **87**(3), 1–33 (2021)
12. 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)