Mining Original Features for Abdominal Organ Segmentation

Zixiao ZHAO Jiahua CHU
AI Innovation and Commercialization Center
National University of Singapore (Suzhou) Research Institute
zixiao.zhao@nusri.cn jiahua.chu@nusri.cn

Abstract

Human abdominal organ segmentation occupies a crucial position in clinical applications. Different organs have rich geometric features, but due to the implicit learning nature of neural networks, extant methods cannot use distinct features to enhance the training in a targeted manner. In this paper we propose a new approach that artificially specifies features, such as spatial-frequency features, contour features, etc., without using complex network structures, effectively improving the generalisability of the model. Specifically, we applied a UNet-like structure integrated the wavelet transform as the backbone to learn spatial-frequency features and a contour learning module to enhance the comprehension of target contour.

1. Introduction

In recent years, deep convolutional neural network-based approaches have been very successful in both natural image and medical image segmenta- tion. U-Net[1], based on encoder-decoder and skip connection structures, has become the baseline for almost all medical image segmentation algorithms. Transunet[2] applies the multi-head self-attention mechanism and multi-layer perceptual structure from transformer to U-Net's encoder. Meanwhile Swin-Transformer[3] replaces the U-shaped convolutional structure entirely with transformer's encoder-decoder. Both achieved state-of-the-art performances on Automatic Cardiac Diagnosis Challenge (ACDC)[4] dataset. However their applicability are somewhat limited because these specific network structures and parameters lead them to obtain even the best results for one dataset, but not for other different modalities or organ lesions.

Compared to natural images, medical images benefit from a more pronounced spatial structure and positional certainty. For example, for CT images of COVID-19, the lesion area clearly shows a large flocculent structure. For cardiac segmentation, each individual heart shows a completely consistent enveloping structure. For multi-organ MRI images within the torso, the location of each organ is absolutely invariant. This feature gave us an idea that instead of modifying the structure of the neural network, whether the performance of the algorithm could be optimized by mining original features of the images themselves. The emergence of nnU-Net[5] has verified this conjecture to a certain extent. nnU-Net has currently become the framework with the strongest generalization capability in medical image segmentation without changing the classical structure of U-Net but putting more attention on data augmentation and pre-processing. However, the limitation of nnU-Net is that it still relies on neural network to implicitly learn the deeper features it considers more important.

To introduce artificially specified features in the learning procedure, we propose the Wavelet UNet with Contour structure to learn spatial-frequency features and object contour features explicitly.

2. Method

The pipeline of the proposed method is depicted in Fig.1.

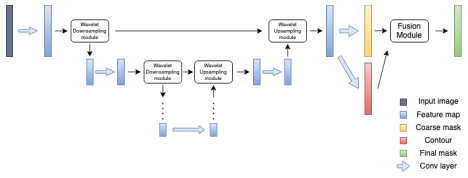


Figure 1: Overview of the proposed solution. Given an image, the network predicts a coarse segmentation mask from wavelet U-Net and a contour prediction, based on which the initial mask is fused and corrected via a feature fusion module to produce the final segmentation mask.

2.1. Preprocessing

The proposed method includes the following preprocessing steps:

- Cropping strategy:
  Due to the limitation of computational resources and

speed requirements, we use LabelSampler inside each image to randomly crop the patches containing the target, the essence of which is that the center pixel of each patch cropped out must be the foreground pixel.

- Intensity normalization method:
  First, the dataset is clipped to the [30.0, 70.0] percentiles of the intensity values of the training dataset. Then a z-score normalization is applied based on the mean and standard deviation of the intensity values.

- Contour extraction:
  To enhance the learning of target contour features, we extracted the contours of each target directly from the annotated mask using connected region analysis and retained their labels.

2.2. Proposed Method

- Network architecture details: A wavelet integrated U-Net[6] is adopted as the base segmentation framework. This modification of U-Net is chosen because for medical images we prefer to learn the structural features characterized in frequency domain.

  Given an input image, the network produces the initial segmentation. Meanwhile, a contour detection module (composed of a simple 1*1 convolution layer) is applied to predict the contour information with the shared features from wavelet U-Net. A feature fusion module is then proposed to balance and combine the initial segmentation mask with the learned contour to generate the final improved segmentation result.

- Loss function: In our architecture, we want the network to optimise three components simultaneously: the result of coarse segmentation, the result of contour reinforcement and the result of final integrated segmentation. Therefore, we calculate the summation of the losses of these three components in the training process, where we use the cross-entropy loss for the coarse segmentation and contour reinforcement, and the Dice Loss for the integrated segmentation result as it has been proved to be robust[7] in medical image segmentation tasks.

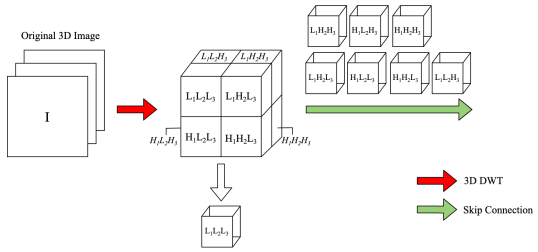
- Number of model parameters: 104874367

- Number of flops: 980053393408

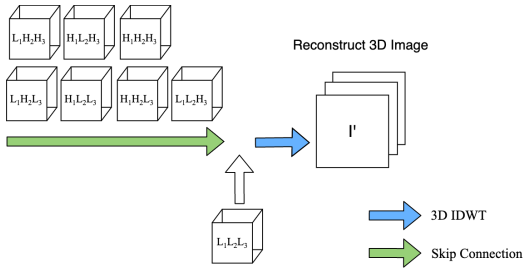
2.3. Post-processing

We resample the output array to the original size and embed it into an output .nii.gz file.

- Resample:
  according to the original voxel size, we resample the output images as its origine.



(a) 3D DWT Module



(b) 3D IDWT Module

Figure 2: The discrete wavelet transform (DWT) is used in the downsampling stage, and the low-frequency part of DWT will be used as the output of the downsampling, while the seven high-frequency parts are retained, fused with the learned low-frequency information after skip connection, and upsampled by the inverse discrete wavelet transform (IDWT).

3. Dataset and Evaluation Metrics

3.1. Dataset

- A short description of the dataset used:
  The dataset used of FLARE2021 is adapted from MSD [8] (Liver [9], Spleen, Pancreas), NIH Pancreas [10, 11, 12], KiTS [13, 14], and Nanjing University under the license permission. For more detail information of the dataset, please refer to the challenge website and [15].

- Details of training / validation / testing splits:
  The total number of cases is 511. An approximate 70%/10%/20% train/validation/testing split is employed resulting in 361 training cases, 50 validation cases, and 100 testing cases. The detail information is presented in Table 1.

3.2. Evaluation Metrics

- Dice Similarity Coefficient (DSC)

- Normalized Surface Distance (NSD)

- Running time

Table 1: Data splits of FLARE2021.

Data Split	Center	Phase	# Num.
Training (361 cases)	The National Institutes of Health Clinical Center	portal venous phase	80
	Memorial Sloan Kettering Cancer Center	portal venous phase	281
Validation (50 cases)	Memorial Sloan Kettering Cancer Center	portal venous phase	5
	University of Minnesota	late arterial phase	25
	7 Medical Centers	various phases	20
Testing (100 cases)	Memorial Sloan Kettering Cancer Center	portal venous phase	5
	University of Minnesota	late arterial phase	25
	7 Medical Centers	various phases	20
	Nanjing University	various phases	50

- Maximum used GPU memory (when the inference is stable)

Table 2: Environments and requirements.

Windows/Ubuntu version	Ubuntu 18.04.4 LTS
CPU	Intel(R) Xeon(R) Gold 6226 CPU @ 2.70GHz
RAM	12×32GB; 2.67MT/s
GPU	8 × Nvidia GeForce RTX 2080Ti
CUDA version	11.1
Programming language	Python3.6.10
Deep learning framework	Pytorch (Torch 1.7.0, torchvision 0.8.0)
Specification of dependencies	torchio 0.18.33
(Optional) code is publicly available at	Currently not publicly available

Table 3: Training protocols.

Data augmentation methods	Random affine, random flip, Gaussian noise.
Initialization of the network	"he" normal initialization
Patch sampling strategy	10 samples in each volume whose center voxel are foreground are randomly chosen as training patches.
Batch size	8
Patch size	128×128×32
Total epochs	500
Optimizer	Adam
Initial learning rate	1e-4
Learning rate decay schedule	poly learning rate policy: $(1 - epoch/10)^{0.8}$
Stopping criteria, and optimal model selection criteria	Stopping criterion is reaching the maximum number of epoch (500).
Training time	83 hours

4. Implementation Details

4.1. Environments and requirements

The environments and requirements of the baseline method is shown in Table 2.

4.2. Training protocols

The training protocols of the baseline method is shown in Table 3.

4.3. Testing protocols

- Pre-processing steps of the network inputs:
  The same strategy is applied as trainging steps.
- Post-processing steps of the network outputs:
  No post-processing step is used.
- Patch aggregation method:
  We used the default aggregation strategy in Torchio[16]. The overlap areas are cropped when aggregating predictions across patches.

5. Results

5.1. Quantitative results for 5-fold cross validation.

Due to time constraints and incomplete model development, we did not perform a five-fold cross-validation

5.2. Quantitative results on validation set.

Table 4 illustrates the results on validation cases. Due to the failure of the model development, we were not able to successfully predict the corresponding positions, and the

analysis of the occurring problems will be analyzed in detail in the last section.

Table 4: Quantitative results on validation set.

Organ	DSC (%)	NSD (%)
Liver	0.0±0.0	0.0±0.02
Kidney	0.0±0.0	0.0±0.0
Spleen	0.0±0.0	0.01±0.08
Pancreas	0.0±0.0	0.0±0.0

5.3. Qualitative results

Due to the failure of the prediction, we did not perform a qualitative analysis of the results.

6. Discussion and Conclusion

First of all, we would like to express our sincere apologies to all the organizers of this challenge. By understanding the dataset as well as the baseline solution, we designed the present novel model structure. In fact, our current model is only in a very preliminary stage. In order to quickly determine the usability of the model, we conducted preliminary tests on some 2D datasets (including 2D extractions of ACDC dataset and CVC-colon dataset), and the results proved that the structure is indeed valid. However, a clear dilemma was encountered in the application of 3D images.

First of all, in image pre-processing and data augmentation, although we tried some naive methods, after studying the source code of nnUNet, we found that our processing of data is relatively insufficient. Also we do not have time to check validity of the pre-processing. The present results are most likely due to the sequence and parameter configuration of the pre-processing algorithm resulting in a change in the image, which weakens the logical relationship between the original image and the learning target.

Although we did not achieve the desired results now, but we are working on the current model and conducting more research to apply this novel structure in a more rational way.

Acknowledgment

The authors of this paper declare that the segmentation method they implemented for participation in the FLARE challenge has not used any pre-trained models nor additional datasets other than those provided by the organizers.

References

[1] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*, 2015, pp. 234–241. 1

[2] J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang, L. Lu, A. L. Yuille, and Y. Zhou, "Transunet: Transformers make strong encoders for medical image segmentation," *arXiv preprint arXiv:2102.04306*, 2021. 1

[3] H. Cao, Y. Wang, J. Chen, D. Jiang, X. Zhang, Q. Tian, and M. Wang, "Swin-unet: Unet-like pure transformer for medical image segmentation," *arXiv preprint arXiv:2105.05537*, 2021. 1

[4] O. Bernard, A. Lalande, C. Zotti, F. Cervenansky, X. Yang, P.-A. Heng, I. Cetin, K. Lekadir, O. Camara, M. A. G. Ballester *et al.*, "Deep learning techniques for automatic mri cardiac multi-structures segmentation and diagnosis: is the problem solved?" *IEEE transactions on medical imaging*, vol. 37, no. 11, pp. 2514–2525, 2018. 1

[5] F. Isensee, P. F. Jaeger, S. A. Kohl, J. Petersen, and K. H. Maier-Hein, "nnu-net: a self-configuring method for deep learning-based biomedical image segmentation," *Nature Methods*, vol. 18, no. 2, pp. 203–211, 2021. 1

[6] Q. Li and L. Shen, "Wavesnet: Wavelet integrated deep networks for image segmentation," *arXiv preprint arXiv:2005.14461*, 2020. 2

[7] J. Ma, J. Chen, M. Ng, R. Huang, Y. Li, C. Li, X. Yang, and A. L. Martel, "Loss odyssey in medical image segmentation," *Medical Image Analysis*, vol. 71, p. 102035, 2021. 2

[8] A. L. Simpson, M. Antonelli, S. Bakas, M. Bilello, K. Farahani, B. Van Ginneken, A. Kopp-Schneider, B. A. Landman, G. Litjens, B. Menze *et al.*, "A large annotated medical image dataset for the development and evaluation of segmentation algorithms," *arXiv preprint arXiv:1902.09063*, 2019. 2

[9] P. Bilic, P. F. Christ, E. Vorontsov, G. Chlebus, H. Chen, Q. Dou, C.-W. Fu, X. Han, P.-A. Heng, J. Hesser *et al.*, "The liver tumor segmentation benchmark (lits)," *arXiv preprint arXiv:1901.04056*, 2019. 2

[10] H. Roth, A. Farag, E. Turkbey, L. Lu, J. Liu, and R. Summers, "Data from pancreas-ct. the cancer imaging archive (2016)." 2

[11] H. R. Roth, L. Lu, A. Farag, H.-C. Shin, J. Liu, E. B. Turkbey, and R. M. Summers, "Deeporgan: Multi-level deep convolutional networks for automated pancreas segmentation," in *International conference on medical image computing and computer-assisted intervention*. Springer, 2015, pp. 556–564. 2

[12] K. Clark, B. Vendt, K. Smith, J. Freymann, J. Kirby, P. Koppel, S. Moore, S. Phillips, D. Maffitt, M. Pringle *et al.*, "The cancer imaging archive (tcia): maintaining and operating a public information repository," *Journal of digital imaging*, vol. 26, no. 6, pp. 1045–1057, 2013. 2

[13] N. Heller, F. Isensee, K. H. Maier-Hein, X. Hou, C. Xie, F. Li, Y. Nan, G. Mu, Z. Lin, M. Han *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*, vol. 67, p. 101821, 2021. 2

[14] N. Heller, S. McSweeney, M. T. Peterson, S. Peterson, J. Rickman, B. Stai, R. Tejpaul, M. Oestreich, P. Blake, J. Rosenberg *et al.*, "An international challenge to use artificial intelligence to define the state-of-the-art in kidney and kidney tumor segmentation in ct imaging." *American Society of Clinical Oncology*, vol. 38, no. 6, pp. 626–626, 2020. 2

[15] J. Ma, Y. Zhang, S. Gu, C. Zhu, C. Ge, Y. Zhang, X. An, C. Wang, Q. Wang, X. Liu, S. Cao, Q. Zhang, S. Liu, Y. Wang, Y. Li, J. He, and X. Yang, "Abdomenct-1k: Is abdominal organ segmentation a solved problem?" *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021. 2

[16] F. Pérez-García, R. Sparks, and S. Ourselin, "Torchio: a python library for efficient loading, preprocessing, augmentation and patch-based sampling of medical images in deep learning," *Computer Methods and Programs in Biomedicine*, p. 106236, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0169260721003102 3