Efficient Multi-Organ Segmentation Using HRNet And OCRNet

Quoc-Cuong Nguyen
University of Information Technology, VNU-HCM
Ho Chi Minh City, Vietnam
18520206@gm.uit.edu.vn

Tien M. Pham
University of Information Technoogy, VNU-HCM
Ho Chi Minh City, Vietnam
tienpmuit@gmail.com

Duc-Lung Vu
University of Information Technoogy, VNU-HCM
Ho Chi Minh City, Vietnam
lungvd@uit.edu.vn

Abstract

The research literature in medical image, driven by the fast growth of deep learning, has proved their robustness while dueling with medical datasets. But the available methods meet difficuties in reaching an efficient speed and memory cost. In this work, we present our method used in the FLARE2021 challenge which aim to solve the requirements of generalization on unseen data and efficiency in terms of computation time and memory cost.

1. Introduction

Segmentation of intra-abdominal organs from medical imaging is an essential step in helping physicians observe and diagnose intra-abdominal injuries. FLARE2021 is the interesting challenge containing realistic challenges of abdominal segmentation problems.

The first challenge comes from the diversity of the dataset, which including multi-center, multi-phase, multi-vendor, and multi-disease cases. Well generalization of the applied method is required.

The second challenge comes from the efficiency requirement for the proposed solutions.

To addressing above difficulties, we employed the 2D segmentation pipeline using OCRNet architecture along with HRNet.

2. Method

To address the challenge of speed and memory, we employed a 2D segmentation model. The input which is in the format of MRI image will be splitted into many slices along its 3rd dimension. The segmentation results of these slices will be stacked to form the output

2.1. Preprocessing

The pixel values are firstly normalized so that they take value between $0 - 255$.

In order to improve model's generalization, we applied some augmentation techniques during the training phase.

- Random crop with crop size 256x256 and padding
- Shift, scale and rotate 2D image.
- Distort current image with several types of transformations. Sequentially, they are applied following this order: randomly change brightness, convert the color format from BGR to HSV, change saturation, change hue, convert back to BGR, change constrast.

2.2. Proposed Method

We compared different encoder and decoder combinations, tried the classic Resnet [1], EfficientNet [2] as the encoder, FCN, Deeplab3+ [3] as the decoder. In the end, we end up choosing the SOTA backbone HRNetV2 [4] and segmentation decoder OCRNet [5] to solve this task. The architecture is illustrated in 1

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

We did a lot of experiments on the loss function, and finally used focal loss [6] to increase the learning weight of the small and thin object classes.

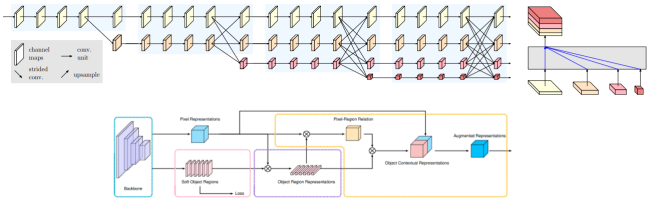

Figure 1. Illustration of the HRNetV2 + OCRNet architecture. The upper half is the main body and representation head of HRNetV2. The lower half illustrates the pipeline of OCR. This figure was taken from the original papers

Our model has a total of 70354170 parameters and its number of flops is 162178878720.

3. Dataset and Evaluation Metrics

3.1. Dataset

- A short description of the dataset used:
  The dataset used of FLARE2021 is adapted from MSD [7] (Liver [8], Spleen, Pancreas), NIH Pancreas [9, 10, 11], KiTS [12, 13], and Nanjing University under the license permission. For more detail information of the dataset, please refer to the challenge website and [14].

- 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

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

4. Implementation Details

4.1. Environments and requirements

We developed our solution using Pytorch as deep learning framework. Nibabel library was used to retrieve the image data from the neuroimaging files. A description of the environment used for deployment of the method, including but not limited to the items illustrated in Table 2.

Table 2. Environments and requirements.

Windows/Ubuntu version	Ubuntu 18.04.5 LTS
CPU	2vCPU of Colab Pro Notebook @ 2.2GHz
RAM	13×4GB; 2.67MT/s
GPU	Nvidia V100
CUDA version	11.0
Programming language	Python3.8
Deep learning framework	Pytorch (Torch 1.9.0, TorchVision 0.10.0)
Specification of dependencies	HRNetV2 + OCRNet
(Optional) code is publicly available at	Code

4.2. Training protocols

We exploit the robustness of HRNetV2 backbone by applying directly the decoder part on the final concatenated feature maps extracted. The decoder part is implemented following an cascaded style. Namely, many decode heads are stacked together, each head takes the output of previous head as its input. In our settings, 2 decode heads, particularly FCN and object-contextual pooling (OCP), are used. Due to the dependence of performance on batch size, we try to increase the batch size to be as large as possible. Here,

by converting the weight value to 16-bit precision, we are able to use the batch size of 8. We employ Adam as our optimizer with a learning rate of 0.001. Full description of the training protocols, including but not limited to the items illustrated in Table 3.

Table 3. Training protocols.

Data augmentation methods	Rotations, scaling, optical distortion
Initialization of the network	He weight initialization
Patch sampling strategy	None
Batch size	8
Patch size	256×256
Total epochs	5
Optimizer	Adam
Initial learning rate	10^{-2}
Learning rate decay schedule	poly learning rate policy: $(1 - epoch/1000)^{0.9}$
Stopping criteria, and optimal model selection criteria	Stopping criterion is reaching the maximum number of epoch + EarlyStopping.
Training time	23 hours
CO_2eq[1]	None

4.3. Testing protocols

After loading the images, the same preprocessing strategy as during training is applied to format the pixel's value. Though the crop size while training model is 256×256, we fed directly the image whose shape is 512×512 into the model to get rid of loss caused by resizing.

5. Results

5.1. Quantitative results on validation set.

Table 4 illustrates the results on validation cases. In term of DSC, the performance on liver segmentation takes high value and low dispersed distributions, which in turn indicate good performance on this organ. The DSC values of kidney and spleen are quite good, too. But their distributions are larger, which means that the performance may be less consistent between different cases. For NSD, the obtained values and the dispersed distributions observed from the violin plots indicate unsatisfying segmentation performance for all four organs.

5.2. Qualitative results

Figure 2 presents some challenging examples. The first row falls into the case of too large upper kidney. As shown below, the model misclassify this area as a combination of spleen and pancrea. The second row presents an example that have its kidney and spleen too close. Here, both of them are classified as a single class which is spleen.

Table 4. Quantitative results on validation set.

Organ	DSC (%)	NSD (%)
Liver	96.27±3.44	80.81±12.10
Kidney	92.71±7.41	82.08±11.44
Spleen	91.81±13.77	82.83±17.25
Pancreas	73.80±17.62	58.30±16.10

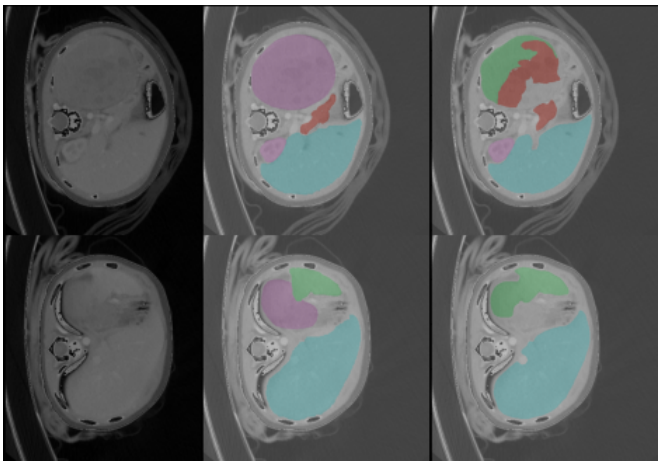

Figure 2. Challenging examples.First column is the image, second column is the ground truth, and third column is the predicted results

6. Discussion and Conclusion

What kind of cases the proposed method works well?

The baseline method can work well on cases where no diseases exist. Besides, the DSC and NSD scores of liver segmentation is higher than the other organs, indicating liver maybe a comparable easier task as a result of its bigger size and consistent shape. Disappointing performance is obtained for pancreas segmentation as a result of the inter-patient anatomical variability of volume and shape.

What are the possible reasons for the failed cases?

The model acts inefficiently in term of segmenting pancrea. A noticeable reason is that this class usually has a sensitive boundary. In case of too close organs, it may also fail because of difficulties in splitting them into different areas.

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] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *CoRR*, vol. abs/1512.03385, 2015. [Online]. Available: http://arxiv.org/abs/1512.03385 1

[2] M. Tan and Q. V. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," *CoRR*, vol. abs/1905.11946, 2019. [Online]. Available: http://arxiv.org/abs/1905.11946 1

[3] L. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," *CoRR*, vol. abs/1706.05587, 2017. [Online]. Available: http://arxiv.org/abs/1706.05587 1

[4] J. Wang, K. Sun, T. Cheng, B. Jiang, C. Deng, Y. Zhao, D. Liu, Y. Mu, M. Tan, X. Wang, W. Liu, and B. Xiao, "Deep high-resolution representation learning for visual recognition," *CoRR*, vol. abs/1908.07919, 2019. [Online]. Available: http://arxiv.org/abs/1908.07919 1

[5] Y. Yuan, X. Chen, and J. Wang, "Object-contextual representations for semantic segmentation," *CoRR*, vol. abs/1909.11065, 2019. [Online]. Available: http://arxiv.org/abs/1909.11065 1

[6] T. Lin, P. Goyal, R. B. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," *CoRR*, vol. abs/1708.02002, 2017. [Online]. Available: http://arxiv.org/abs/1708.02002 2

[7] 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

[8] 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

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

[10] 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

[11] 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

[12] 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

[13] 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

[14] 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