3D Unet-based Kidney and Kidney Tumer Segmentation with Attentive Feature Learning

Qingyi Tao, Zhonghua Wu, Isaac Cheong, Jingyi Yang, Zongyuan Ge, Guosheng Lin, Jianfei Cai

No Institute Given

Abstract. To study the kidney diseases and kidney tumor from Computed Tomography(CT) imaging data, it is helpful to segment the region of interest through computer aided auto-segmentation tool. In the KiTs 2019 challenge [1], we are provided 3D volumetric CT data to train a model for kidney and kidney tumor segmentation. We introduce an improved deep 3D Unet by enriching the feature representation in CT images using an attention module. We achieve 1.5% improvement in the segmentation accuracy when evaluated on the validation set.

1 Introduction

In our solution, we used a generic 3D U-Net to train a multi-label segmentation model. In particular, we adopt nnU-Net framework [2] as the base framework. nnU-Net is proposed for Medical Imaging Decathlon as a general framework for various organs and tumors in CT and MRI data. However, it has not been tested on the kidney and kidney tumor. It is designed to adapt to many different medical imaging datasets through automatic data analysis and preprocessing. On top of this baseline, given that the kidney tumours tend to only take up a relatively small portion of the CT slice and that they also appear to be rather difficult to distinguish from the kidney itself, we follow the approach in [3] to apply a spatial attention mechanism on the generic 3D U-Net to improve the feature representation.

2 Approach

Preprocessing: We use the interpolated version of raw data provided by the official website in order to use 3D CNN. Same as in the original nnU-Net module, the data are normalized based on the training set, subtracted by the mean value and divided by the standard deviation. The volumes are clipped by keeping only the values within 0.05-0.95 percentile.

Attentive 3D U-Net: We illustrate the brief structure of our network in 1. To improve the feature representation, we add an attention module similar to [3] at the middle layer. It contains a convolution layer (ϕ_{att}) followed by a softmax function and a max normalization operation. The process can be mathematically written as:

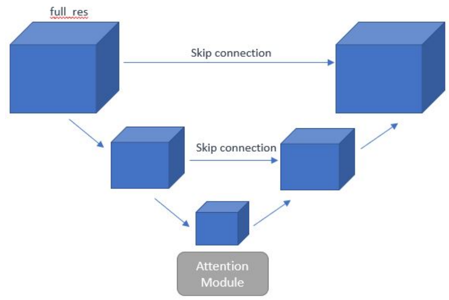

Fig. 1. Attentive 3D U-Net.

$$att_i = \phi_{att}(X_i)/\tau \tag{1}$$

$$att_{i,d}^{'w,h} = \frac{exp(att_{i,d}^{w,h})}{\sum_{w=1}^{att}\sum_{h=1}^{H} exp(att_{i,d}^{w,h})} \tag{2}$$

$$att_{i,d}^{''w,h} = \frac{att_{i,d}^{'w,h}}{\max\limits_{w,h} |att_{i,d}^{'w,h}|} \tag{3}$$

$$X_i^{out} = att_i^{''} \otimes att_i. \tag{4}$$

where X_i is the ith input feature map, (w, h) is the shape of the feature map, and d is the channel index. $\otimes$ represents the element-wise multiplication operation. τ is the temperature of the softmax operation and it is set to 3 in our implementation.

Training of 3D Unet: We follow the training procedures as described by [2]. We modify the orignal nnU-Net training to a parallel training that could be run on 4 GPUs concurrently. The batch size is set to be 8.

References

1. Heller, N., Sathianathen, N., Kalapara, A., Walczak, E., Moore, K., Kaluzniak, H., Rosenberg, J., Blake, P., Rengel, Z., Oestreich, M., Dean, J., Tradewell, M., Shah, A., Tejpaul, R., Edgerton, Z., Peterson, M., Raza, S., Regmi, S., Papanikolopoulos, N., Weight, C.: The kits19 challenge data: 300 kidney tumor cases with clinical context, ct semantic segmentations, and surgical outcomes (2019)
2. Isensee, F., Petersen, J., Klein, A., Zimmerer, D., Jaeger, P.F., Kohl, S., Wasserthal, J., Koehler, G., Norajitra, T., Wirkert, S., et al.: nnu-net: Self-adapting framework for u-net-based medical image segmentation. arXiv preprint arXiv:1809.10486 (2018)
3. Tao, Q., Ge, Z., Cai, J., Yin, J., See, S.: Improving deep lesion detection using 3d contextual and spatial attention. arXiv preprint arXiv:1907.04052 (2019)