Kidney and Tumor Segmentation Based on 3D Context Extracting

Bin Yan

Beijing Laboratory of Intelligent Information Technology
School of Computer Science, Beijing Institute of Technology, Beijing, 100081, P.R. China
2120171086@bit.edu.cn

Abstract. Organ segmentation and lesion detection play a vital role in the computer-aided diagnosis (CAD) systems. The task of this Kits challenge is about kidney and tumor segmentation. We proposed an effective model to complete this Kits challenge. Our model receives part of body 3D scans as input, and outputs the probability map of the input scans. 2D contexts of intra-slices are extracted by VGG network, and 3D contexts of inter-slices are presented by concatenating the 2D contexts. Then proposals are extracted by region proposal network (RPN), while 3D context are regarded as auxiliary information for region of interest (ROI) regression, classification and mask generation. Our model has shown promising result for this Kits challenge.

Keywords: Organ Segmentation · Kidney · Tumor.

1 Introduction

Medical image analysis based on deep learning has gradually been a hot research topic. It can reduce the workload of doctors and make their work more efficient.

Currently, methods for medical image segmentation can be roughly divided into four categories: 2D fully convolutional network (FCN) [1], 3D convolutional network methods [5], trip-planar schemes [6], and recurrent neural network (RNN) methods [2]. The 2D convolutional network methods are usually applied to each slice to obtain 2D segmentation, such as U-Net [7], DCAN [1], etc. The disadvantage of this kind of methods is that the 3D contexts are not in consideration, resulting in a lot of information between the slices being ignored. Therefore, many medical image segmentation methods [2, 5] based on 3D convolutional networks and RNN have been proposed. The methods based on RNN use the LSTM [4] which take the feature maps of the slices as input, and extract the relationship between the slices which is also called 3D context, while the feature maps of the slices are extracted by the 2D convolutional network. The 3D convolutional network methods take the overall 3D image as input, 3D contexts and 2D contexts are both taken into consideration through the 3D convolutional network. Generally, due to the capacity limitation of the graphics card, the 3D images are often divided into multiple patches as inputs, such as V-Net [5], coarse-to-Fine [9].

We proposed an effective model to complete this Kits challenge. Our model receives part of body 3D scans as input, and outputs the probability map of the input. 3D contexts are considered as auxiliary information for segmentation in our model, specifically, the 3D contexts indicates the anatomical information of the input. Our model has shown promising results, which demonstrates the effectiveness of our model.

2 Dataset and Preprocessing

Dataset The training data set has a total of 210 computed tomography (CT) images. They are all stored at 'imaging.nii.gz' in the NIFTI format. The shape of most CT images is $T \times 512 \times 512$, where T is the number of slices along the z-axis. It is necessary to adjust the shape of all CT images to the common shape $T \times 512 \times 512$. There are three categories of annotations: background, kidney, and tumor, corresponding to the value 0, 1, and 2 respectively, which stored at 'segmentation.nii.gz'. We randomly selected data and divided them into training set, validation set, and test set, each for 60%, 20%, and 20% respectively.

Preprocessing

Resampling Different samples have different slice spacing and pixel spacing, this means that the volume represented by each voxel is different. During the scan process, different sampling slice spacing and pixel spacing lead to different shape for the same human body parts. For example, the sampling slice spacing and pixel spacing are represented by a triple (slice spacing, width spacing, height spacing), for a certain part of body, A and B represents CT images obtained from scan process with different sampling parameters, (1, 1, 1) and (5, 1, 1) respectively. Obviously, a longitudinal compression has occurred and the organs of B appear to be flatter than the organs of A. Such a difference has a great influence on the final segmentation. So first of all, resampling is required to adjust slice spacing and pixel spacing to the same case by interpolation.

Intensity normalization In order to include more target pixels within the smaller window width, we set the window width to [-200, 300] which can contain 99.6% of the target pixels after truncating the raw data, and then linearly change the raw data to [0, 255]. This can make the image contrast relatively higher, making it easier to distinguish between target and non-target organs. We slice the results after normalization and each slice is saved in 'bmp' format.

Mask and Bounding Box We extract the masks and bounding boxes (b-boxes) slice by slice to get the masks and b-boxes of the kidney and tumor in each slice. The masks are used as ground truth to calculate loss and Dice score, and b-boxes are used to get proposals. Each mask and b-box class is also extracted at the same time of course. Thus, for each target organ, there is a ground truth triple (mask, b-box, class), and according to the Kits challenge rules, only (mask, class) will be outputted and saved in a prescribed format.

Augmentation In order to avoid over-fitting as much as possible, we performed a simple data augmentation to augment the data set. We adopted mirroring, scaling, and Gaussian blurring strategies to increase the diversity of data sets.

3 Method

In general, there is a strong prior knowledge of body anatomical structure, and the organs are roughly fixed in a certain body part. So it is of great significance to make use of the prior knowledge of human body anatomical prior to segment organs. Previous researches [2, 9] have confirmed that the anatomical prior can effectively improve the accuracy of organ segmentation, so we first extract 3D context information as the anatomical prior for segmentation.

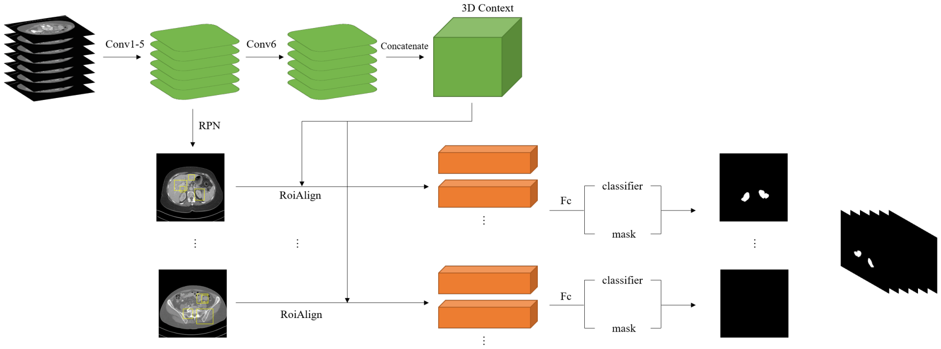

Fig. 1. Our model employs the VGG network to extract context of slice, and all the contexts are regarded as 3D context after concatenating. ROI align is employed in 3D context to extract features, then followed classification and mask functions. The output is probability maps and corresponding class of each ROI area, which then is transformed into masks in a prescribed format. Finally, concatenating all the masks as the final segmentation.

Our model is shown in Fig. 1. Motivated by 3DCE model [8], the convolutional blocks (Conv1-5 of VGG-16) are employed to each slice for extracting the feature map, and then the feature map of each slice and the corresponding bounding boxes and classes are seeded into the region proposal network (RPN) to obtain the proposals of the slice. Then the feature map of each slice is inputted into another convolutional layer (Conv6 of VGG-16). Then the feature maps are concatenated to aggregate the 3D context of the CT image. While in the process of input, three consecutive slices (s-1, s, s+1) are regarded as an image to the VGG network to extract the feature map, where s is the current target slice.

Similar with Mask R-CNN [3], we then take the regions proposals and 3D context as input and send them into the ROI align layer to obtain the predicted ROIs, while Mask R-CNN takes the 2D pyramid features as input instead. Finally, the predicted ROIs and 3D context are regarded as input and are sent into the classification function and the mask function respectively, and the classes and masks of the ROI area are obtained at the end.

Roi Align 3DCE model[8] adopts ROI pooling layer, which ignores the symmetry of the spatial structure because of the nearest neighbor interpolation. We adopt ROI align layer, which has been shown its effectiveness by Mask R-CNN [3]. ROI align replaces the nearest neighbor interpolation by bilinear interpolation, thus preserving the symmetry of structure.

Classification The classification function receives the ROIs and 3D contexts as input, and the output is a one-dimensional vector of length N, where N is the number of categories of organs, and $N = 2$ in this Kits challenge. The structure of the classification is a convolution layer with 7 × 7 kernel size, followed by a batch-norm layer. Then followed a convolution layer with 1 × 1 kernel size and a batch-norm layer, the ReLU active layer and a linear layer with an output dimension of N. The final is the softmax layer. Since bounding boxes are not need for this Kits challenge, which is different from the lesion detection task, we remove the linear layer from the original classification function in Mask R-CNN [3], which is used for bounding boxes regression.

Mask Same with the classification, the mask function takes both ROIs and 3D contexts as input and outputs a probability map of the same shape as the target slice s. The structure of mask function is a padding layer which kernel size is 3 × 3, and followed by three convolution layers and all their kernel size is 3 × 3, each followed by a batch-norm layer, a deconvolution layer which kernel size is 2 × 2 and stride is 2. The last convolution layer with kernel size of 1 × 1 is used to adjust the output dimensions. Of course, there is still sigmoid layer and ReLU layer at the end.

Loss We adopt the standard DICE coefficient as our loss function.

$$s = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{2}\left(\frac{2 * n_{t,tp}^{(i)}}{2 * n_{t,tp}^{(i)} + n_{t,fp}^{(i)} + n_{t,fn}^{(i)}} + \frac{2 * n_{k,tp}^{(i)}}{2 * n_{k,tp}^{(i)} + n_{k,fp}^{(i)} + n_{k,fn}^{(i)}}\right)$$

where N is the number of categories of organs, n presents for the class label at each voxel.

4 Implementation

During the training process, due to the capacity limitation of the GPU, we take 48 slices of the complete CT image in the z-axis direction as the input, and the resolution of each slice is scaled to 256 × 256. We trained the model on two GPUs, which type is GTX 1080 Ti with 11 GB memory. And some of our experimental results are shown in Fig. 2.

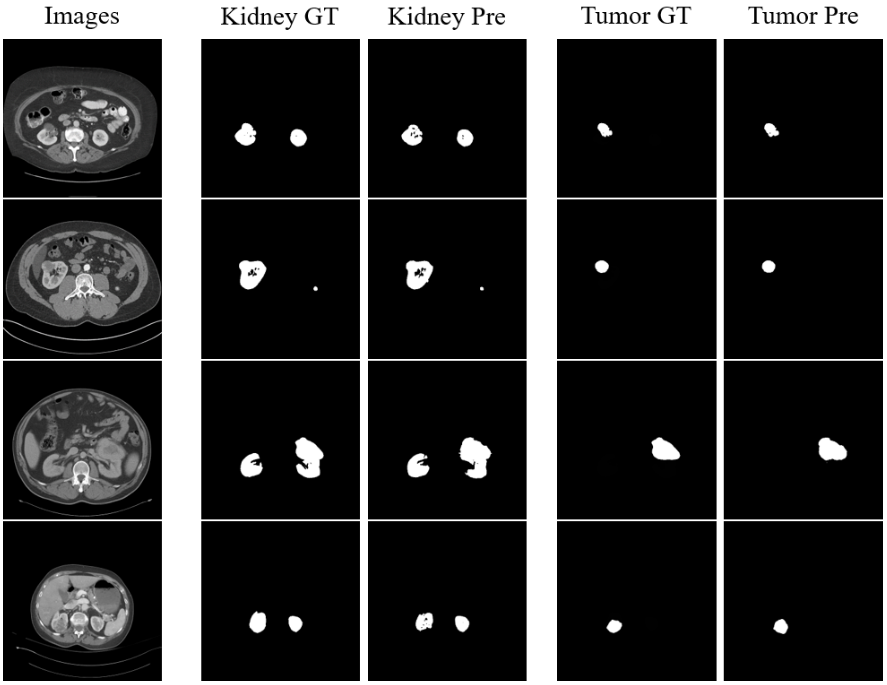

Fig. 2. The results of our model. From left to right: sources images of cases scans, corresponding kidney ground truth, kidney prediction, tumor ground truth and tumor prediction.

5 Conclusions

In this Kits challenge, we proposed a segmentation model base on 3D context extracting. By taking 3D context in consideration, our model can obtain the accurate kidney and tumor segmentation. In the future work, we will optimize the method of 3D context extracting and make it more generalizable for other segmentation task.

References

1. Chen, H., Qi, X., Yu, L., Heng, P.: DCAN: deep contour-aware networks for accurate gland segmentation. CoRR **abs/1604.02677** (2016), http://arxiv.org/abs/1604.02677

2. Chen, J., Yang, L., Zhang, Y., Alber, M., Chen, D.Z.: Combining fully convolutional and recurrent neural networks for 3d biomedical image segmentation. In: Lee, D.D., Sugiyama, M., Luxburg, U.V., Guyon, I., Garnett, R. (eds.) Advances in Neural Information Processing Systems 29, pp. 3036–3044. Curran Associates, Inc. (2016), http://papers.nips.cc/paper/6448-combining-fully-convolutional-and-recurrent-neural-networks-for-3d-biomedical-image-segmentation.pdf
3. He, K., Gkioxari, G., Dollár, P., Girshick, R.B.: Mask R-CNN. CoRR **abs/1703.06870** (2017), http://arxiv.org/abs/1703.06870
4. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Computation **9**(8), 1735–1780 (1997). https://doi.org/10.1162/neco.1997.9.8.1735, https://doi.org/10.1162/neco.1997.9.8.1735
5. Milletari, F., Navab, N., Ahmadi, S.: V-net: Fully convolutional neural networks for volumetric medical image segmentation. CoRR **abs/1606.04797** (2016), http://arxiv.org/abs/1606.04797
6. Prasoon, A., Petersen, K., Igel, C., Lauze, F., Dam, E., Nielsen, M.: Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network. In: Mori, K., Sakuma, I., Sato, Y., Barillot, C., Navab, N. (eds.) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2013. pp. 246–253. Springer Berlin Heidelberg, Berlin, Heidelberg (2013)
7. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. CoRR **abs/1505.04597** (2015), http://arxiv.org/abs/1505.04597
8. Yan, K., Bagheri, M., Summers, R.M.: 3d context enhanced region-based convolutional neural network for end-to-end lesion detection. In: MICCAI (2018)
9. Zhu, Z., Xia, Y., Shen, W., Fishman, E.K., Yuille, A.L.: A 3d coarse-to-fine framework for automatic pancreas segmentation. CoRR **abs/1712.00201** (2017), http://arxiv.org/abs/1712.00201