MULTI-ORGAN SEGMENTATION USING SIMPLIFIED DENSE V-NET WITH POST PROCESSING

Ming Feng, Weiquan Huang, Yin Wang, Yuxia Xie

Tongji University, Shanghai, China
{1810865, 1730784, yinw, yuxia_xie}@tongji.edu.cn

ABSTRACT

With the recent advances in the field of computer vision, Convolutional Neural Networks (CNNs) are widely used in organ segmentation of computed tomography (CT) images. Based on the Dense V-net model, this paper proposes a simplified version with postprocessing methods to help reduce the fragments in organ segmentation results. Comparing with the baseline method that uses a sharpmask model with conditional random field (SM+CRF), our model improves the Dice ratio of Esophagus, Heart, Trachea, and Aorta by 10%, 4%, 7%, and 6%, respectively.

***Index Terms*—** Convolutional Neural Networks, CT Segmentation, Dense V-net

1. INTRODUCTION

Organ segmentation of CT images is of great importance in medical diagnosis. The identification and localization of organs are the daily work of the radiologist. Since CT images are complex and three-dimensional(3D), distinguishing organs manually is a difficult and tedious task. Therefore, segmentation using deep learning methods automatically have received a great deal of attention in medical imaging research. In the field of 3D medical image segmentation, there are two main methods. The first is to segment each slice independently, e.g., using the U-net model [1]. The other is to use the 3D convolution to aggregate inter-slice information and to segment all slices of the CT image at once, e.g., V-net [2] is one of the 3D convolutional network models for this purpose. Gibson et al. [3] integrated the two-dimensional segmentation model of Dense net [4] into V-net and proposed a Dense V-net architecture for multiple organ segmentation. Overall, single slice segmentation methods cannot utilize inter-layer dependencies for better results but are computationally more efficient. All slice 3D segmentation can aggregate all layers for better accuracy but is more expensive to compute.

In this paper, we present our multi-organ segmentation solution used in the SegTHOR challenge hosted at the ISBI'19 conference. Observing that the training data is relatively small and easy to overfit deep convolutional neural nets, we simplify the Dense V-net model to achieve better results with the testing data. Our postprocessing method further reduces fragments in the prediction mask. The overall improvement over the SM+CRF baseline model [5] is between 4 to 10 percent over different organs.

2. OUR MODEL

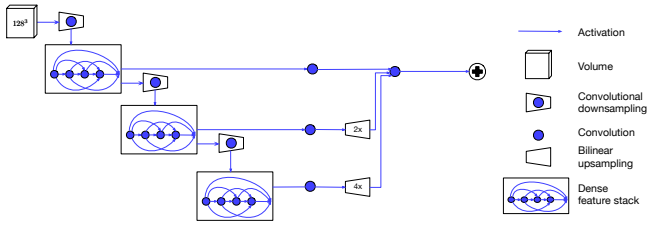


Fig. 1. Simplified Dense V-net model

The structure of our proposed model is shown in Fig. 1. Comparing with the original Dense V-net model, there are two main differences. First, the input size is different. The input size of the original model is 144^3. The number of partial data slices in our data is less than 144, so we set the input size to 128^3. Second, the spatial prior block is discarded.

The encoder block of the segmentation network generates three sets of feature maps of different sizes. The decoder block upsamples the smaller feature maps so the output mask is of the same size as the input image. The output layer generates the segmentation mask with the probability vector of different segmentation classes at each pixel.

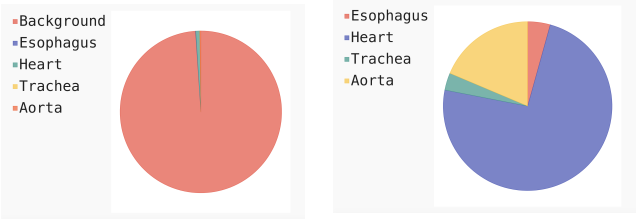
3. IMPLEMENTATION

This section discusses various optimization techniques to reduce the Dice loss and to minimize the Hausdorff distance.

3.1. Data preprocessing

Preprocessing is part of our fully automated organ segmentation method. By analyzing the training data provided, we find the following issues.

First, the dataset is small and is quite easy to overfit our deep neural networks. Second, for a single CT slice, the proportion of pixels of various organs is quite different. Fig. 2 shows the imbalance of different organs at different slices. Last, considering the relative position of the machine and the person while scanning, the CT images can be scaled and rotated. Based on these observations, we apply the following techniques.



(a) The ratio of background to organs.

(b) The ratio of organs.

Fig. 2. Background and organ volume proportion in training data.

3.1.1. Patch sampling

We ensure that each class is sampled with the same probability. According to the slice range of the test dataset, the sample block size is set to 128^3.

3.1.2. Data augmentation

During the training stage, we randomly rotate pictures (within -10°−10°) and randomly scale pictures (-10 %−10 % range). We implement the data augmentation on the Niftynet framework [6]. The data augmentation method used in the training stage will not affect the structure of the Dense V-net.

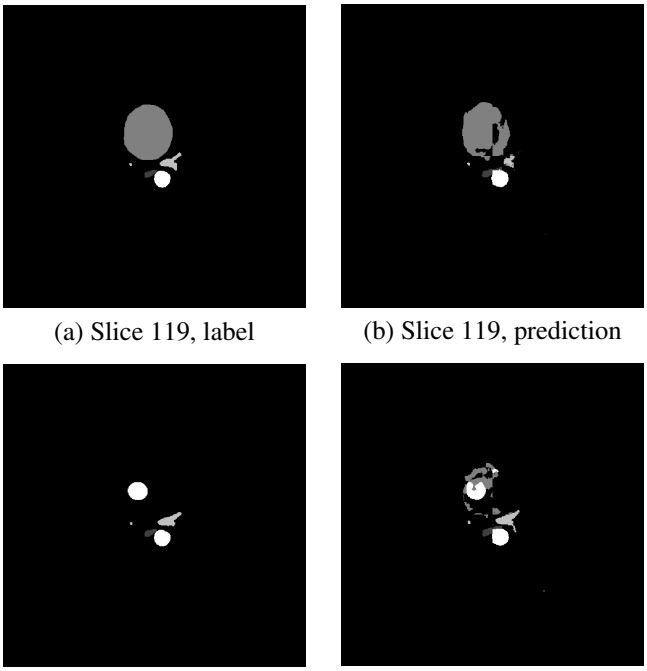
3.2. Postprocessing

By comparing the prediction result with the ground truth label, we find the following issues.

In the training data organs are all connected, but organs are not connected in the predicted results. Some areas of the CT image are not smooth, Fig.3. There are multiple organ inclusions in the same slice, which does not actually exist. In the prediction result, the organ is connected but there are background noise inside.

For the first question above, we experimented with the following methods.

The CT image is sliced along three dimensions respectively, then count the number of connected blocks of each organ. For each dimension and each organ, the largest connected block is retained, and the other parts are considered background noise and therefore removed. Experiments show that our method achieves obvious increase; see Algorithm 1.



(a) Slice 119, label

(b) Slice 119, prediction

(c) Slice 120, label

(d) Slice 120, prediction

Fig. 3. The 119th and 120th slices of patient 30 in the labeled data and prediction result. We can see that the heart disappears at the 120th slice in the labeled data. The sudden disappearing of an organ often leads to incorrect predictions.

Fig.4 shows the predict results with the removal of disconnected blocks. The CT image is sliced along the depth direction, for each layer, 5^2 average filtering is used, which seriously affects the segmentation results of small sample organs like Esophagus and Trachea, and has little effect on multi-sample organs as Heart and Aorta. The enlargement of the organs for each class within each layer has little effect on the segmentation.

3.3. DicePlusXEnt loss function

The loss functions commonly used in segmentation are Cross-Entropy loss and Dice loss. The Cross-Entropy loss examines each pixel separately, and compares the prediction results with one-hot encoded target vector. It does not consider the imbalance of different segmentation classes, and can lead to poor prediction results with the minority classes. Imbalanced classes are very common in medical image segmentation. The Dice loss is essentially a measurement of the overlap between the predicted mask and the ground truth mask, calculated as follows [7] :

$$l_{dice} = -\frac{2}{|K|}\sum_{k\in K}\frac{\sum_{i\in I} u_i^k v_i^k}{\sum_{i\in I} u_i^k + \sum_{i\in I} v_i^k} \quad (1)$$

where K is the set of segmentation classes, I is the entire

Algorithm 1 Axis-based denoise method

Input: The result from model, T_m;
Output: Remove noise block prediction result, Q_m;

for all $axis_i$ of T_m **do**
  for all $slice_j$ of the $axis_i$ **do**
    for all $category_k$ of T_m **do**
      Sets $slice[-1]$ and $slice[max+1]$ to -1;
      if The current slice contain the $category_k$ and the previous slice does not contain $category_k$ **then**
        The current slice index is added to $blockIn$;
      end if
      if The current slice contain the $category_k$ and the next layer does not contain $category_k$ **then**
        The current slice index is added to $blockOut$;
      end if
    end for
  end for
  The $blockIn$ corresponds to the $blockOut$ element one by one, each set of them represents a continuous block, the data difference represents the contiguous block length, the contiguous block of the maximum length is reserved, and the other continuous blocks in Q_m are set as the background class;
end for
return Q_m;

image, and u_i^k, v_i^k are the predicted and ground truth value of class k at pixel i, respectively. Dice loss is more suitable for sample's extremely imbalance situation, but in our experience, using the Dice loss alone will adversely affect back propagation, making training extremely unstable.

We use DicePlusXEnt loss [8], which is the sum of the Cross-Entropy loss and the Dice loss, as follows:

$$l_{total} = l_{dice} + l_{CE} \tag{2}$$

This loss function will improve the sample imbalance to a certain extent and improve the stability of network training.

Due to the imbalance of the samples, we set the weight of the Cross-Entropy loss in DicePlusXEnt: $w(Background)$=1, $w(Heart)$=2, $w(Trachea)$=3, $w(Aorta)$=4, $w(Esophagus)$ =5.

4. EXPERIMENTS

Our experiment is conducted on the SegTHOR dataset [5]. Niftynet is used in our model training, which is implemented by Tensorflow. Based on the preprocessed data, the Dense V-net network is trained and then fine-tuned with different parameter configurations.

The activation function used in the network is Leaky ReLU. The batch size is four. We use the Adam optimizer with an initial learning rate of 0.01. If the loss value does

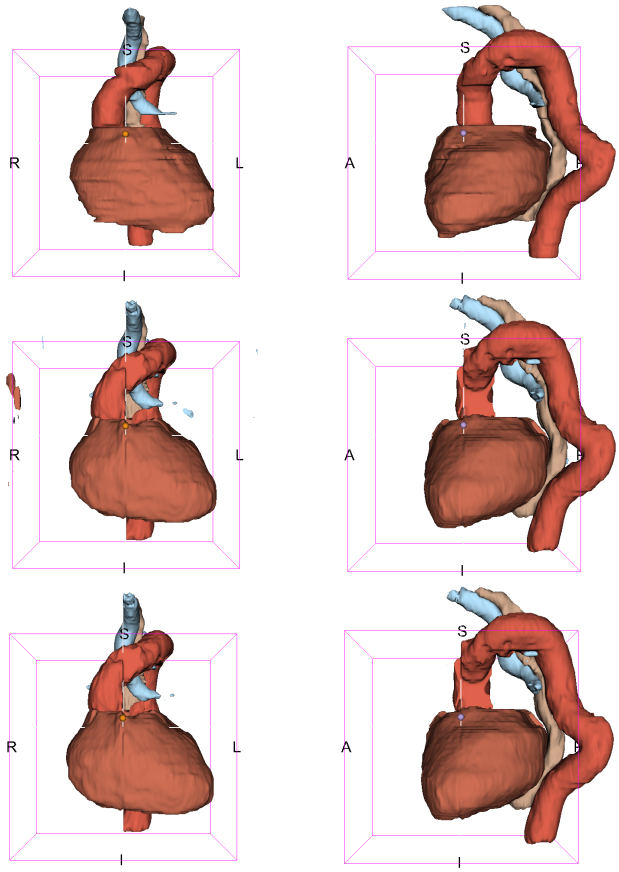


Fig. 4. From top to bottom, main view and the left view of the true label, the predicted result, the 3D denoise. The small fragments are significantly reduced.

Algorithm 2 Training model

Input: The training data, X and label, Y;The fusion model numbers, N;The learning rate list, L;
Output: Segmentation result, R;

for all n_i in $range(N)$ **do**
  for all $l_i \in L$ **do**
    while loss does decrease in 500 iterations **do**
      Forward and backward;
    end while
  end for
  Save the model with the lowest validation set loss during this iteration;
end for
Fusion saved models, get R_{ori};
$R \leftarrow$ Axis-based denoise(R_{ori});
return R;

Table 1. Performance of different methods

Methods	Dice				Hausdorff			
	Esophagus	Heart	Trachea	Aorta	Esophagus	Heart	Trachea	Aorta
Dense V-net (resize sampling)	0.588862	0.906035	0.772924	0.780659	1.531403	0.598427	1.783999	0.997311
Dense V-net (balanced sampling)	0.746470	0.937633	0.875301	0.914082	1.153503	0.221647	1.726525	0.402991
Dense V-net (balanced sampling and average filter)	0.490914	0.914966	0.589199	0.840300	3.246483	0.292705	2.417643	1.066558
Dense V-net (balanced sampling and organ enlargement)	0.486919	0.913697	0.575745	0.841042	4.128935	0.817668	5.587061	1.581914
7 Dense V-net fusion	0.763881	0.940254	0.883234	0.915550	0.771958	0.188203	0.597479	0.308775
7 Dense V-net fusion (1D denoise)	0.763973	0.940255	0.885504	0.915673	0.766507	0.188183	0.330171	0.295968
7 Dense V-net fusion (3D denoise)	0.765423	0.940225	0.885614	0.915954	0.661974	0.188183	0.325847	0.258024
7 Dense V-net fusion (3D denoise and weighted loss)	**0.773450**	**0.941403**	**0.892730**	**0.923325**	**0.640093**	**0.182138**	**0.307711**	**0.235788**

* This Dense V-net is simplified Dense V-net.

not decrease after 500 iterations, then the learning rate decreases by ten-fold, up to 0.0001. When the learning rate is 0.0001 and after 500 iterations if the loss does not change, the learning rate is reset to 0.1. This process is repeated seven times, and the model with the lowest validation loss during the training process is selected for comparison. In addition, we pick the parameters of the minimum loss of the validation set in each training cycle, seven models in total, and fuse the results together for comparison [9]; see Algorithm 2. Table 1 shows the results with different settings.

Overall, the fusion results are much better than the single-model prediction. Denoise in postprocessing further improves the accuracy. Heart and Aorta have much better segmentation results than Esophagus and Trachea.

5. CONCLUSION

Based on the analysis of the training data, we simplified Dense V-net to perform multi-organ segmentation effectively. We use a variety of optimization techniques such as multi-scale prediction, data augmentation, and data postprocessing to improve the stability and performance of the model. Comparing to the baseline model of SM+CRF [5], the Dice rate of organ segmentation is improved up to 10%. After our optimization, there is still room for improvement for small organs, and delineation algorithms could help to refine organ boundaries.

6. REFERENCES

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

[2] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *2016 Fourth International Conference on 3D Vision (3DV)*. IEEE, 2016, pp. 565–571.

[3] Eli Gibson, Francesco Giganti, Yipeng Hu, Ester Bonmati, Steve Bandula, Kurinchi Gurusamy, Brian Davidson, Stephen P Pereira, Matthew J Clarkson, and Dean C Barratt, "Automatic multi-organ segmentation on abdominal ct with dense v-networks," *IEEE transactions on medical imaging*, vol. 37, no. 8, pp. 1822–1834, 2018.

[4] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.

[5] Roger Trullo, Caroline Petitjean, Su Ruan, Bernard Dubray, D Nie, and D Shen, "Segmentation of organs at risk in thoracic ct images using a sharpmask architecture and conditional random fields," in *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*. IEEE, 2017, pp. 1003–1006.

[6] Eli Gibson, Wenqi Li, Carole Sudre, Lucas Fidon, Dzhoshkun I Shakir, Guotai Wang, Zach Eaton-Rosen, Robert Gray, Tom Doel, Yipeng Hu, et al., "Niftynet: a deep-learning platform for medical imaging," *Computer methods and programs in biomedicine*, vol. 158, pp. 113–122, 2018.

[7] Carole H Sudre, Wenqi Li, Tom Vercauteren, Sebastien Ourselin, and M Jorge Cardoso, "Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations," in *Deep learning in medical image analysis and multimodal learning for clinical decision support*, pp. 240–248. Springer, 2017.

[8] Fabian Isensee, Jens Petersen, Andre Klein, David Zimmerer, Paul F Jaeger, Simon Kohl, Jakob Wasserthal, Gregor Koehler, Tobias Norajitra, Sebastian Wirkert, et al., "nnu-net: Self-adapting framework for u-net-based medical image segmentation," *arXiv preprint arXiv:1809.10486*, 2018.

[9] Leslie N Smith, "Cyclical learning rates for training neural networks," in *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, 2017, pp. 464–472.