A 3D Convolutional Neural Network Framework for Polyp Candidates Detection on the Limited Dataset of CT Colonography

Yizhi Chen, Yacheng Ren, Ling Fu, Junfeng Xiong, Rasmus Larsson, Xiaowei Xu, Jianqi Sun and Jun Zhao* *Member, IEEE*

***Abstract*—Proper training of convolutional neural networks (CNNs) requires annotated training datasets of large size, which are not currently available in CT colonography (CTC). In this paper, we propose a well-designed framework to address the challenging problem of data shortage in the training of 3D CNN for the detection of polyp candidates, which is the first and crucial part of the computer-aided diagnosis (CAD) of CTC. Our scheme relies on the following two aspects to reduce overfitting: 1) mass data augmentation, and 2) a flat 3D residual fully convolutional network (FCN). In the first aspect, we utilize extensive rotation, translation, and scaling with *continuous* value to provide numerous data samples. In the second aspect, we adapt the well-known V-Net to a flat residual FCN to resolve the problem of detection other than segmentation. Our proposed framework does not rely on accurate colon segmentation nor any electrical cleansing of tagged fluid, and experimental results show that it can still achieve high sensitivity with much fewer false positives. Code has been made available at: http://github.com/chenyzstju/ctc_screening_cnn.**

I. Introduction

Colorectal cancer, also known as colon cancer, is becoming an urgent threat to the health of human life. Colorectal cancer is now the third most commonly diagnosed cancer in the United States with 135,430 new cases estimated in 2017 [1]. And it is also the second leading cause of cancer death with 50,260 cases estimated in 2017 [1]. Computed tomography colonography (CTC) is a minimally invasive CT-based screening technique and has been endorsed for colorectal cancer screening [2].

In the past decades, dozens of computer-aided diagnosis (CAD)-CTC systems have been proposed. A typical CAD-CTC system consists of two stages, candidate detection and false-positive (FP) reduction. Although most of the CAD systems are effective in yielding a high detection sensitivity, they tend to generate a number of false-positive detections mainly because of the inefficiency of the method of candidate detection. Besides, most of the existing CAD-CTC systems are designed with delicate manually designed image features, which heavily rely on the experience of experts and often fail to handle the balance between complexity and generalization.

In the recent years, deep convolutional neural networks (DCNN) have rapidly become a powerful choice of algorithm for the analysis of medical images. The applications of convolutional neural networks (CNNs) range from classification, detection, segmentation, registration to other tasks in medical imaging. And the scope of target anatomical organs includes brains, eyes, chests, breasts, hearts and so on [3]. However, in the field of CAD-CTC, the application of CNN suffers from the data shortage of available annotated datasets. It is well known that typical CNN systems have millions of parameters, whose proper training require a large number of annotated datasets. Generally, the lack of annotated data samples would cause severe overfitting problem. 3D medical images, other than 2D ones, make the overfitting problem even worse.

Until now there have been only a very few CNN applications in CT colonography.

Nappi et al. proposed a CNN system utilizing transfer learning to reduce the number of false positives of the output of traditional CAD systems [4]. Umehara et al. further designed an ensemble of CNN systems utilizing the same technique of transfer learning [5]. Though effective, they make compromises by degenerating the dimension of detection problems from three into two and heavily rely on the output of traditional classic CAD systems.

Roth et al. proposed a CNN scheme of random view aggregation for FP reduction, which is good but still degenerates into 2.5D and generates polyp candidates in a traditional way [6].

In this study we focus on designing a 3D fully convolutional network (FCN) scheme for the first yet crucial part of the CAD-CTC system, namely, candidate detection. Although our proposed framework only concentrates on the candidate detection problem, the overall performance of CAD systems can be improved if the number of FPs produced by the stage of candidate detection can be efficiently reduced. And there is no doubt that the experience of application of 3D CNN to candidate detection can be generalized to the other part of CAD systems.

In order to tackle the tough problem of data shortage in 3D space, we propose a mass data augmentation technique to effectively enlarge the number of data samples and a flat 3D FCN with residual learning to reduce the number of parameters. Both parts are equally necessary to address the problem of overfitting.

This study is partly supported by National Key R&D Program of China (2016YFC0104608), the National Natural Science Foundation of China (grant No. U1732119, grant No. 11575115, No. 81371634), and Shanghai Jiao Tong University Medical Engineering Cross Research Funds (YG2017ZD10, and YG2014ZD05). *An asterisk indicates the corresponding author.*

Yizhi Chen, Yacheng Ren, Ling Fu, Junfeng Xiong, Rasmus Larsson, Xiaowei Xu and Jianqi Sun are with School of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai, 200240, China

*Jun Zhao is with School of Biomedical Engineering, SJTU-UIH Institute for Medical Imaging Technology, and MED-X Research Institute, of Shanghai Jiao Tong University, Shanghai, 200240, China. (email: junzhao@sjtu.edu.cn).

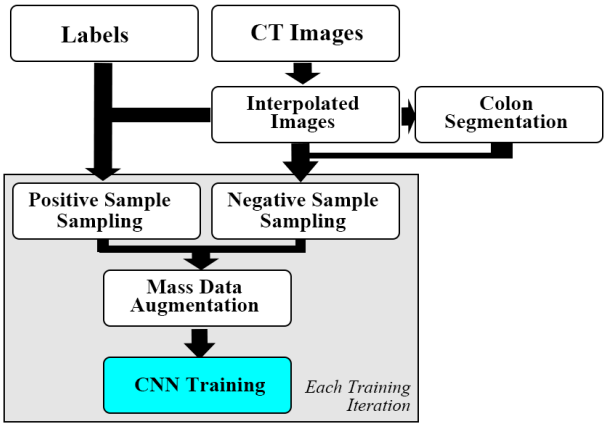

Figure 1. Block diagram of the overall CNN training scheme

II. MATERIALS

144 CTC patient cases are sampled from CTC collection of TCIA [7-9], and CTC collection of Dr. Richard Choi, Virtual Colonoscopy Center, Walter Reed Army Medical Center (WRAMC).[1] All patients underwent bowel preparation with the oral administration of contrast agents and insufflation of air. Because each patient case includes two scans, supine and prone, there are in total 288 CT images with 386 confirmed polyps. 260 polyps have a diameter larger than or equal to 5mm and smaller than 10mm, and 126 polyps have a diameter larger than or equal to 10mm. All the confirmed polyps in the CT images have been correlated with the colonoscopy finds by experienced radiologists.

III. METHODS

A. Overview

The overall structure of the training of our proposed system is shown in Fig. 1. Firstly, the original CT images are interpolated into isotropic images by linear interpolation in the axial direction. Secondly, the colon lumen is segmented from the isotropic CT images [10]. The colon segments are only used to select inputting areas for the following training. Therefore, the segmentation of colon walls does not have to be precise as long as the colon wall is included in the segmentation. Thirdly, cropped CT patches and their corresponding labels of equal size are sampled uniformly from the colon lumen area. Data Augmentation is applied during the sampling procedure. Lastly, data samples are used to train the proposed flat residual FCN in an end-to-end way.

B. Sample Generation

Cube-like samples with a fixed length are drawn from the isotropic interpolated CT images.

The sampling of negative samples matters a lot since the number of positive samples is quite limited. In other words, although there are not enough positive samples for CNNs to learn what is a polyp, we can alternatively offer networks opportunities to learn what is not a polyp. In each training iteration, polyps in the training set are randomly selected as positive samples, while non-polyp areas are randomly and uniformly sampled as negative samples.

[1] The protocol was approved by the review board of the corresponding institutions.

C. Mass Data Augmentation

Unlike common computer vision problem, medical images are delicate and full of noise. Thus, the transformation used should be under certain limit to train a reliable and reasonable model. The data augmentation we apply is a combination of several linear transformation, including rotation, translation, zooming and flipping.

Let α be the length of sides of the input cube of the network. Clearly α should be larger than the diameter of the biggest polyp in the training dataset. Let τ be the diameter of the polyp in the input sample. Let β be the diameter of the smallest polyp in the training dataset.

The procedure of data augmentation is in the following order:

- Zooming. The size of a polyp is a vital indicator to the determination of its malignant significance, but is irrelevant to the detection of itself. Thus, polyp data can be safely zoomed bigger or smaller to produce samples of different size, making the network scale-invariant. Zooming rates with **continuous random values** in a wide range are applied here. Remember that the following inequality must be satisfied to make data sample reasonable:

$$\lambda\tau < \alpha \quad (1)$$

$$\lambda\tau > \beta \quad (2)$$

- Rotation. Because the normal vector angles of polyps do not affect the CAD diagnosis，rotation can be safely applied. Considering the large amount of the computation it would take to resample an 3D image, the angle of rotation transformation is regularly taken as in the form of regular discrete sparse [0º, 90º, 180º, 270º]. However, **continuous random values** in both degrees of freedom (DOF) are taken here to enable a fully free 3D rotation. Rather than producing only 16 combinations, free rotation could generate more augmentations from one sample. The new permutation factor T is:

$$\mathrm{T} = \left(\frac{2\pi}{\tan^{-1} 2/\alpha}\right)^2 \quad (3)$$

  T is estimated under the assumption that the smallest rotation is to move the farthest point of input from the rotation center in a distance of one pixel. Given an α of 48, multiple will be about 25000. In fact, this is underestimated as rotation can take any continuous value.

- Translation. Translation is the one of the easiest way of data augmentation, but it is of equal importance. Translation of larger range is applied as:

$$\left[-\frac{\alpha}{4}, \frac{\alpha}{4}\right] \quad (4)$$

- Flipping. Regular flipping is applied.

Compared to the common augmentation setting [11], **a much bigger range** is applied in rotation, zooming, and translation, and **continuous value** is taken. When combined,

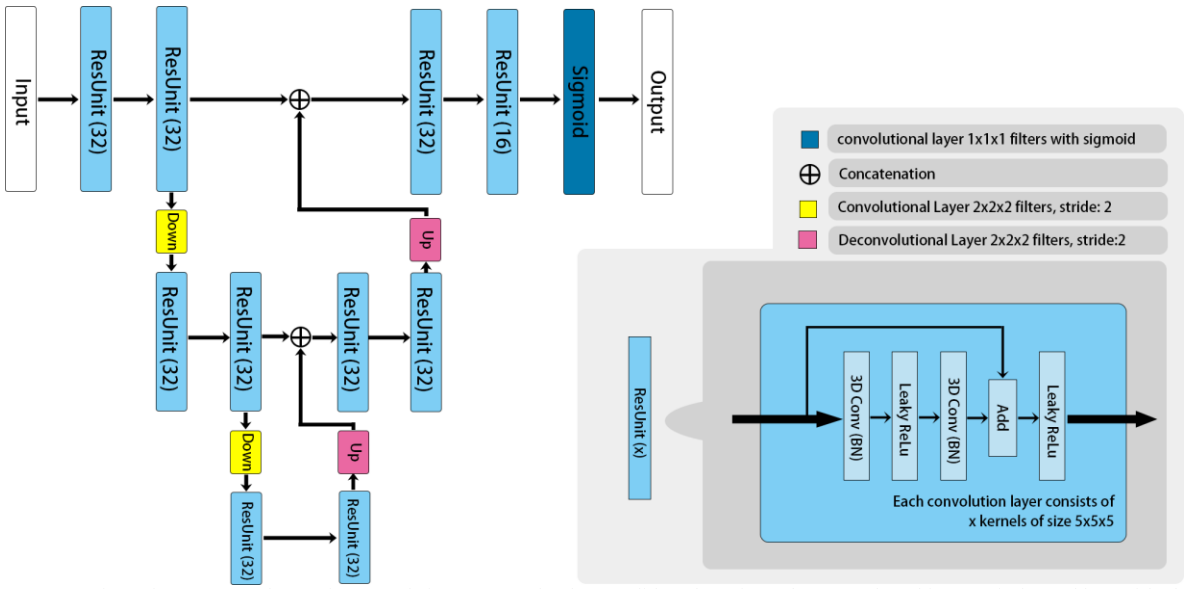

Figure 2. Overview of our proposed FCN framework for automated polyp candidate detection. The network architecture is formed by residual units, which are demonstrated in the lower right corner

our method is able to produce billions of samples from just one image. The augmented images are all different in density, and thus the network has to capture the higher semantic features, avoiding being stuck in simple low-level density features.

D. Flat Residual FCN

In Fig. 2 we provide a schematic representation of our proposed FCN. The network is mainly made up of three stages together with two compression components and two corresponding decompression components. Therefore, the network is able to catch various image patterns from high semantic level to low statistic level.

In each stage, common convolutional layers are replaced with residual units. It is known that CNNs learn better when stacking more layers. But the vanishing/exploding gradients make it rather troublesome to train a very deep neural network. As [12] indicates, residual functions can help networks go deeper while still keeping them simple and slim.

The overall structure of our proposed FCN is similar to the V-Net [13]. The main difference lies in the aims of the networks: V-Net is for image segmentation while the proposed FCN here is for detection. Therefore, the size of CAD-CTC input is much smaller, while the complexity of the problem is about the same, if not higher. Because of the smaller size of the input, the number of down convolution layers we can apply here is quite limited. To compensate for the reduction of stages, more residual units are introduced in each stage to keep the neural network deep, which is essential in training a smart network. These changes result in a flat residual FCN optimized for detection problems.

The number of channels for each convolutional layer are also restricted to control the scale of the parameters.

The network is trained with a loss function based on Dice coefficient as in [13]. L2 loss regularization on kernels is also applied.

IV. Experiment and Results

The proposed framework was evaluated and validated on the two data collections mentioned above by a five-fold cross validation. The evaluation metrics are sensitivity and average number of FPs per scan. A detection is counted as true positive if it overlaps any annotated polyps.

A. Implementation Details

According to our dataset, α was set to 48. It is a relatively big number in a screening scenario, but only by this can the largest polyp in the training dataset be fit into the input to the network. Similarly, β was set to 5.

The network was randomly initialized from a Gaussian Distribution N(0, 0.05), and the learning rate was initialized as 0.001 with an exponential decay of 0.85 every 1000 batches. In the test time, the exponential moving average of parameters was applied to provide a more stable and robust prediction.

The method was implemented in the Tensorflow framework using three NVIDIA 1080ti GPUs. One GPU was used to calculate the real-time mass data augmentation, while

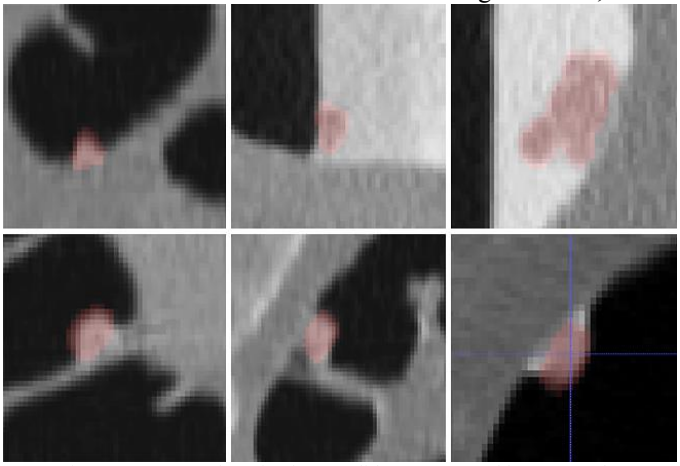

Figure 3. Examples of polyp candidates generated by CNN. The segmentations of detected polyps are viewed here as the red masks.

TABLE 1. COMPARISIONS OF ALGORITHMS

Methods	Years	Polyps	Polyps size(mm)	Sensitivity	FPs/scan
Roth et al. [6]	2015	252	≥5	Almost 1	147.4
Ren et al. [14]	2017	197	≥5	1.00	58.8
Proposed	2018	386	≥5	0.97	20.3

the other two were used to load and train CNN models in parallel to acquire a speed boost.

B. Detection Result and Comparison with Other Methods

The proposed method achieves a sensitivity of 97% with 20 FPs per dataset, and outstanding segmentation performance with an average dice ratio of 80% considering the various shapes of polyps. Fig. 3 shows examples generated by the trained CNN.

Note that there does not exist any public CT colonography dataset with annotated polyps. Thus our proposed framework cannot be directly compared with previous published algorithms, since they are tested on different datasets. In Table 1, a relative comparison of different polyp screening scheme is made. Three traditional screening methods recently published are compared.

Compared to methods utilizing low-level image features like shape index and curvedness [6, 14], our proposed method produces much fewer FPs while keeping a very high sensitivity. The proposed method is believed to be more robust and stable since it is trained end to end and does not rely on precise colon segmentation nor any electrical colon cleansing, both of which are more likely to cause trouble.

C. Ablation Study

To further investigate the contribution of the mass data augmentation, an ablation study was carried out. Capability of polyp screening is evaluated using the same FCN under two different configurations. One with the proposed mass data augmentation, and the other with regular flipping, random rotation of [0º, 90º, 180º, 270º] in all the two DOFs, and normal translation in a range of 4 pixels. As Figure 4 shows, the proposed method stands advantageous.

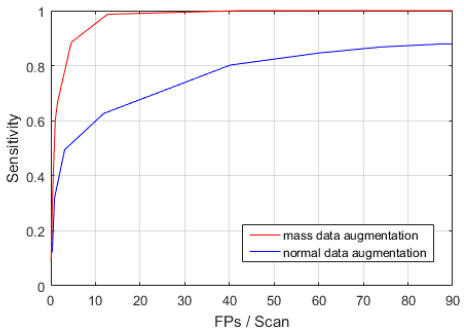

Figure 4. FROC curves of the performance of CNN under different data augmentation settings

V. DISCUSSION AND CONCLUSION

In this paper a polyp candidate generation framework based on deep learning was proposed. Compared to previous classic methods, it gains more efficiency and robustness by directly learning from the data in 3D space. To overcome the lack of annotated data in the application of CNNs to CTC, we employ mass data augmentation and advanced network structure to make what was impossible before now possible. For CTC, we believe that a fully 3D CNN scheme, covering both candidate detection and FP reduction, will be a better solution to polyp detection in the coming future.

REFERENCES

[1] American Cancer Society, "Cancer Facts & Figures 2017," *Atlanta: American Cancer Society*, 2017.

[2] P. Maisonneuve, E. Botteri, and A. B. Lowenfels, "Screening and surveillance for the early detection of colorectal cancer and adenomatous polyps," *Gastroenterology,* vol. 135, no. 2, pp. 710, 2008.

[3] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. van der Laak, B. van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *arXiv preprint arXiv:1702.05747*, 2017.

[4] G. D. Tourassi, S. G. Armato, J. J. Näppi, T. Hironaka, D. Regge, and H. Yoshida, "Deep transfer learning of virtual endoluminal views for the detection of polyps in CT colonography," vol. 9785, pp. 97852B, 2016.

[5] K. Umehara, J. J. Näppi, T. Hironaka, D. Regge, T. Ishida, and H. Yoshida, "Deep ensemble learning of virtual endoluminal views for polyp detection in CT colonography." pp. 101340G-101340G-6.

[6] H. R. Roth, L. Lu, J. Liu, J. Yao, A. Seff, K. Cherry, L. Kim, and R. M. Summers, "Improving Computer-Aided Detection Using Convolutional Neural Networks and Random View Aggregation," *IEEE Trans Med Imaging,* vol. 35, no. 5, pp. 1170-81, May, 2016.

[7] K. Smith, K. Clark, W. Bennett, T. Nolan, J. Kirby, M. Wolfsberger, J. Moulton, B. Vendt, J. Freymann, "Data From CT_COLONOGRAPHY," 2015.

[8] C. D. Johnson, M.-H. Chen, A. Y. Toledano, J. P. Heiken, A. Dachman, M. D. Kuo, C. O. Menias, B. Siewert, J. I. Cheema, and R. G. Obregon, "Accuracy of CT colonography for detection of large adenomas and cancers," *New England Journal of Medicine,* vol. 359, no. 12, pp. 1207-1217, 2008.

[9] K. Clark, B. Vendt, K. Smith, J. Freymann, J. Kirby, P. Koppel, S. Moore, S. Phillips, D. Maffitt, M. Pringle, L. Tarbox, F. Prior, "The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository," *Journal of digital imaging,* vol. 26, no. 6, pp. 1045-1057, 2013.

[10] L. Lu, D. Zhang, L. Li, and J. Zhao, "Fully automated colon segmentation for the computation of complete colon centerline in virtual colonoscopy," *IEEE Trans Biomed Eng,* vol. 59, no. 4, pp. 996-1004, Apr, 2012.

[11] Q. Dou, H. Chen, Y. Jin, H. Lin, J. Qin, and P.-A. Heng, "Automated Pulmonary Nodule Detection via 3D ConvNets with Online Sample Filtering and Hybrid-Loss Residual Learning." pp. 630-638.

[12] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition." *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770-778, 2016.

[13] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation." *3D Vision (3DV), 2016 Fourth International Conference on.* IEEE, pp. 565-571, 2016

[14] Y. Ren, J. Ma, J. Xiong, L. Lu, and J. Zhao, "High-Performance CAD-CTC Scheme Using Shape Index, Multiscale Enhancement Filters, and Radiomic Features," *IEEE Trans Biomed Eng,* vol. 64, no. 8, pp. 1924-1934, Aug, 2017.