Towards One-Stage Framework: Optimization of 3D FCNs for Polyp Detection in CT Colonography

Yizhi Chen[a], Puming Zhang[a], Jun Zhao[a, b, *]

[a] School of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai, 200240, China

[b] SJTU-UIH Institute for Medical Imaging Institute, and MED-X Research Institute of Shanghai Jiao Tong University, Shanghai, 200240, China

[*] Corresponding author. E-mail address: junzhao@sjtu.edu.cn. Postal address: Wexuan Building 309, Dongchuan Road 800, Shanghai, China. 13621706907

Abstract

Purpose: Typical computer-aided diagnosis (CAD) systems for computed tomography colonography (CTC) consist of two crucial steps, including candidate screening and false positive (FP) reduction. However, simpler solutions with equal or even better performance are always preferred.
Methods: A powerful one-stage framework for polyp detection utilizing the fully convolutional network (FCN) technique is proposed. The stage of FP reduction, the electrical cleansing for the contrast-enhanced fluid and the accurate colon wall modeling are now no longer needed. The root of our work lies in the adaptation of segmentation-specialized FCNs for the task of polyp detection, which mainly comprises three components: 1) a multistep training scheme, 2) an ensemble loss function, and 3) adaptive data transform. The proposed network can be directly trained with the pixel-level polyp segmentation labels, and thus, information on the size, shape and position of the polyps can be seamlessly integrated. Inadequacy of the labeled data, which has long been a problem for the application of 3D CNNs in the field of CAD-CTC, can be largely alleviated.
Results: The proposed method was evaluated using 547 oral contrast-enhanced CT scans with 390 polyps and achieved high by-polyp sensitivity with a low FP rate for both large polyps (>=10 mm) and small polyps (<10 mm). With 1.5 FPs per scan, the by-polyp sensitivity was 99.1%/95.3% for large/small polyps.
Conclusion: The experiment results demonstrate that our proposed framework obtained a superior performance especially on small polyps compared with several state-of-the-art methods.

Keywords

Deep learning; CT colonography; fully convolutional network; computer-aided detection;

1. Introduction

It has been decades since colorectal cancer become one of the deadliest cancers. According to the American Cancer Society [1], colorectal cancer is now the second leading cause of cancer-related mortality in the United States, with 50630 cases estimated in 2018, and the fourth most commonly diagnosed cancer, with 140150 cases estimated. Since most malignant colorectal tumors arise from colonic polyps in a transformation process of up to 15 years, the risk of colon cancer can be significantly reduced by the early detection of polyps [2]. Various prototypes have been developed for the detection and diagnosis of colon polyps, among which computer tomography colonography (CTC) is a minimally invasive, feasible clinical alternative. It offers a three-dimensional (3D) whole-belly view and frees the patients from the pain brought by the invasive colonoscopy techniques.

Dozens of computer-aided diagnosis (CAD) systems have been proposed since the 2000s to eliminate the efforts that doctors have to take to go through hundreds of CT slices to catch a tiny polyp.

1.1 Non-CNN Methods

In the past two decades, many CAD systems, based on geometric feature extraction and classical statistical classification and clustering models, have demonstrated their efficacy

For polyp candidate detection, the main approaches can be separated into several categories, such as the surface-geometry-based methods and the gradient based methods. The surface-geometry-based methods include colon wall thickness [3], surface curvature[4], sphere fitting[5] and so on. The gradient-based method includes shape index and curvedness [6], surface normal overlap (SNO) [7] and so on. Other methods include colon wall analysis [3, 8].

For the FP reduction stage, the published methods can also be separated into several categories, including supervised classification, feature enhancement [9] and the supine-prone registration approaches [10, 11].The supervised classification methods include the feature extraction stage and the classification stage. Feature extraction methods include the Edge Displacement Field-Based method [12], random orthogonal shape section (ROSS) [13], second principal curvature flow [14], multiscale enhancement filters and radiomic features [15, 16], combination of 3D Laplacian of the Gaussian filter and 2D projection image features [17], position-based geometric and texture features [18]. The classification methods include LDA (linear discriminant analysis) [6], Mahalanobis distance based linear classifier [12], support vector machine (SVM) [13], random forest [15, 16] and so on.

Generally, the efficacy of the CAD systems introduced above resides in the complex image processing pipeline, consisting of the colon surface segmentation, electrical colon cleansing, polyp candidate screening, feature extraction and FP reduction.

Most of the existing CAD systems relies heavily on the precise segmentation of the colon surface and the correct localization and vivid restoration of the contrast-enhanced fluid. Although a number of papers have been published to tackle the problems [19-22], the complicated fickle nature of the colon makes it extremely hard to explicitly model every situation.

1.2 CNN Methods

Recently, attempts have been dedicated to learning features in a data-driven way in computer vision, among which the convolutional neural network (CNN) has been the most distinguished and promising method. Due to their end-to-end nature and robust adaptation capability, CNNs have enjoyed a great success in the once challenging natural image recognition tasks. Enormous progress has been made in the domain of image classification [23], object detection [24-26], and semantic segmentation. It should be emphasized that CNN is also rapidly gaining a dominating advantage in the automated diagnosis of medical images, including lung nodule detection [27], image synthesizing [28], and prostate segmentation [29].

Until now, there have been only several CNN applications in CT colonography. A deep-learning-based multispectral electronic cleansing approach was previously introduced [30]. For polyp detection,

transfer learning on 2D rendering images of polyps was applied to reduce the number of FPs generated by the previous traditional CAD systems [31], and an ensemble of these was further assembled to enhance the system [32]. To take advantage of the high-dimensional contextual information, a 2.5-D framework was complemented, with a random view aggregation for FP reduction [33]. The methods above still rely on the output of the traditional methods of polyp candidate screening and probably suffer from the loss of spatial information because of the degeneration of the data dimension from the original 3D space.

1.3 Contributions

A straightforward approach of employing CNN techniques would be replacing the classic two-step framework with two different CNNs. However, considering the uncertainty brought on by the increased model complexity and the external training workload, a simpler solution is always preferred. In the paper we present a straightforward one-stage CAD framework utilizing 3D CCN technique. In contrast to other common two-step CAD frameworks with candidate screening and FP reduction, our system consists of only one step and can be trained end-to-end.

The main module of our detection scheme is a 3D fully convolutional network (FCN) with an encode-decode U-Net structure. The one-stage framework is made possible by an FCN with elaborate optimization for the polyp detection problem to address the challenge in the complex 3D space of medical CT volumes. As part of the optimization, an adaptive data transform, a multistep training scheme and an ensemble loss function were employed to enhance the network capability so that the detection problem could be handled by only one neural network.

Our main contributions are:

1. Rather than the classic two-step CAD scheme, a simple CAD-CTC framework composed of only one polyp detection step is proposed. With elaborate training settings optimized for the polyp detection challenge, FP reduction is no longer needed.
2. Successful adaptation of a segmentation-specialized FCN for the task of polyp detection in the domain of medical images is achieved. Considering that the tasks of object segmentation and object detection are closely related, several techniques are applied together to improve the detection capability of the FCN.
3. Unlike previously published CAD-CTC systems with 2D or 2.5D CNNs, the proposed framework utilizes a 3D CNN technique. Training data are adaptively augmented in 3D space for the more efficient data-driven learning.

2 Materials

Data tested in this paper were acquired from two open datasets, namely, the CTC collection of TCIA [34-36] and the CTC collection of the Water Reed Army Medical Center (Dr. Richard Choi, Virtual Colonoscopy Center). Insufflation of air before CT scanning and bowel preparation with an oral administration of contrast agents were applied to the patients for a better visualization of the colon. The current of the CT scans was 200-390 mA, with 120 kVp voltage. Each patient case contained two CT scans from the supine and prone positions. An example is shown in Fig. 1.

In this study, 277 patients were randomly selected from the CTC datasets and were labeled at the pixel level by two specialists according to the provided ground-truth polyp location information. 145 of the patients were diagnosed with one or more polyps. Since a patient had two different scans, there were a total of 547 CT scans, and 281 of them had one or more than one polyp.

If the same polyp was from two different scans, it was considered one polyp, and thus, there were 205 polyps in total, with 64 polyps larger than or equal to 10 mm and 141 polyps smaller than 10 mm. On the other hand, if the same polyp from different two scans was considered a different polyp, there were 390 polyps in total, with 126 polyps bigger than or equal to 10 mm and 264 polyps smaller than 10 mm. The shape distribution and the size distribution are shown in Table 1 and Fig. 2. Of the 390 individual polyps, 145 of them are located on the colon fold (37.18%). And 109 of them are totally or half submerged in contrast-enhanced fluid (27.95%).

Table 1. *Shape distribution of the polyps used.*

Shape	Sessile	Pedunculated	Round	Flat	Oval	Not recorded
Number	99 (48.3%)	34 (16.6%)	34 (6.8%)	14 (2.4%)	2 (1.0%)	51 (24.9%)

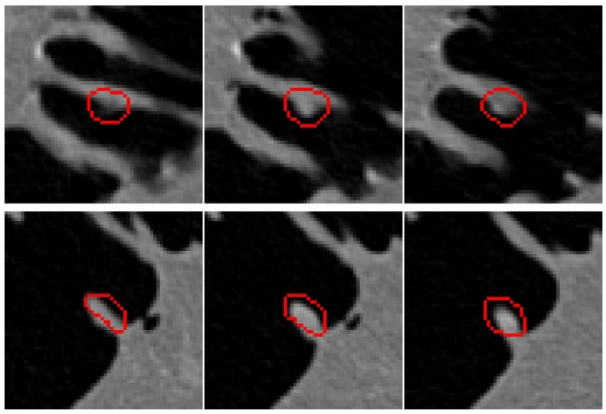

Fig. 1. Figures of a polyp example. The two rows represent the same polyp from two corresponding CT scans. Each row has three slices of the same polyp outlined by red lines. The lower one itself is very hard to detect if the other CT scan is neglected.

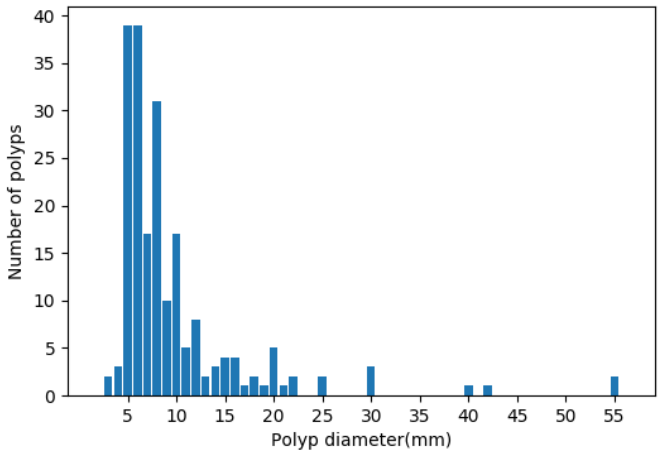

Fig. 2. Size Distribution of the polyps.

3 Methodology

3.1. Overview

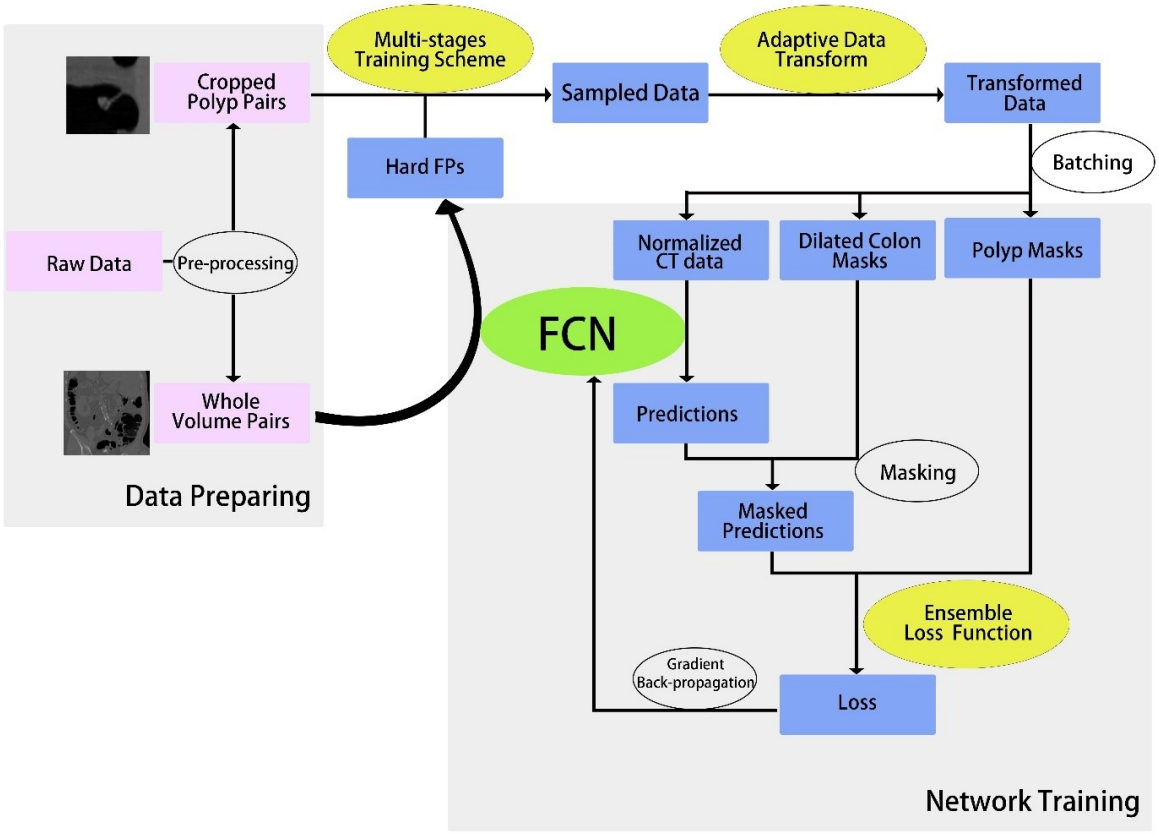

Fig. 3. Training workflow of the proposed method. The image data are under elaborate sampling and transformation before being fed into the neural networks for training.

A colon segmentation algorithm was utilized to generate the inner colon lumen [20], and dilation was applied to include all the surrounding area. Then a trained FCN was applied on the located colon area, and the output score map of the FCN was under threshold to produce polyp predictions. The maximum score in all the voxels of a polyp prediction was considered as its significance level.

As is stated, no explicit colon wall modeling is required as long as the colon segmentation algorithm is able to locate the colon area. Therefore, the bar for the required segmentation algorithm is low. Accurate colon segmentation and electrical cleansing for tagged materials were not prerequisites for our proposed polyp detection system.

The FCN is the core module in our proposed polyp detection system, and an overview of its training scheme is illustrated in Fig. 3. First, an isotropic dataset was generated from the polyp database, including two types of data pairs. Then cubic data blocks were sampled and cropped from the dataset and under an adaptive free 3D linear transform. The cubic data blocks were accumulated into batches and sent into the FCN to obtain the polyp predictions. The predictions were masked by dilated colon

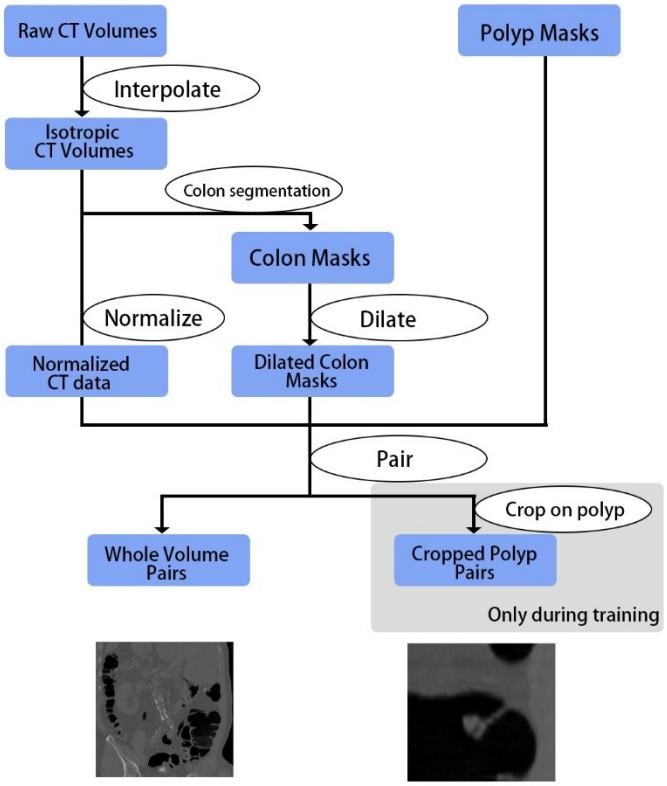

Fig. 4. Diagram of the preprocessing module.

masks to focus on the colon area. Lastly, the FCN was trained on the calculated ensemble loss. The trained FCN was deployed to find the hard FPs, which were sent back into the training dataset to boost the training efficiency.

In the inference time, the trained FCN was applied on the CT volume by the overlapped sliding window technique. Note that the sliding length in the test time is different than that in the train time. A bigger length of the sliced window, denoted as l_{test_window}, was applied to accelerate the inference. The overlap length for the sliding window method is denoted as $l_{test\_overlap}$.

3.2 Data Preprocessing

The data preprocessing procedure is elucidated in Fig. 4. Since the proposed CAD system works in a 3D space and requires a free 3D linear transform, linear interpolation was applied to ensure the isotropy of data. Then, CT data were linearly normalized to [0, 1]

Two kinds of training data were generated. Firstly, the normalized CT data were paired with the corresponding dilated colon mask and polyp label mask, resulting in an instance of whole volume pair. The second ones were cropped polyp pairs. For each labeled polyp, smaller cubic volumes with a length of l_{polyp_volume} were cropped in the polyp centers. The point of generating two types of volume data is later discussed in section 3.4.1.

3.3 FCN Inference

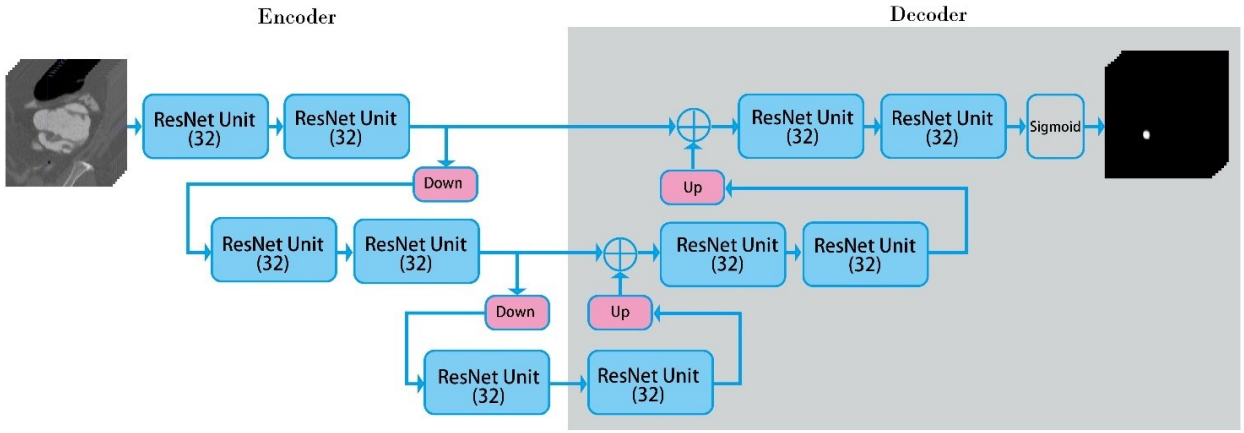

Fig. 5. Illustration of the proposed FCN. The rectangle, denoted by the ResNet Unit, is the basic unit, as in Fig. 6. The one denoted by Down is the convolutional layer, with a stride of 2, acting like a pooling layer. The one denoted by Up is the deconvolutional layer. The one denoted by Sigmoid is a sigmoid layer to restrict the output of the network. The ⊕ represents the concatenation of channels.

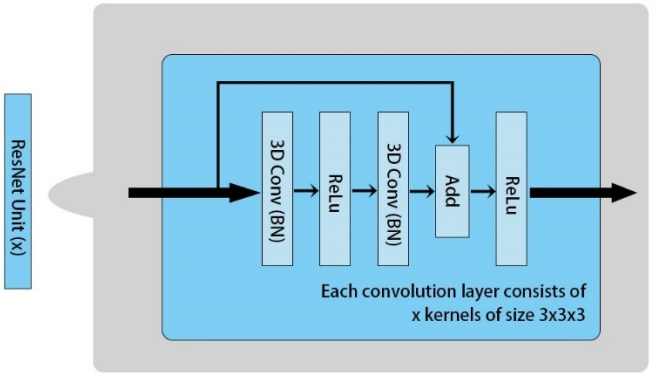

Fig. 6. Illustration of a basic ResNet unit. It contains two 3D convolutional layers, which are each followed by a batch normalization layer. The x in the ResNet Unit(x) represents the number of the channels in the convolutional operators.

Considering that the polyps in our dataset were labeled at the voxel level, we utilized an FCN as the detection module of our framework. The network structure is illustrated in Fig. 5 and Fig. 6. Because FCNs are capable of retaining spatial information, additional information of shape, position and size for polyps can be naturally integrated into network training. FCNs are trained with volume outputs, which are equivalent to groups of classification examples, and thus it is reasonable to apply FCNs to the limited-dataset-size polyp detection problem.

3D kernels were utilized to construct our model. It is now well known that 3D CNNs are capable of encoding richer contextual representations and extracting more discriminative spatial features. However, due to the quickly increasing number of parameters, 3D CNNs are harder to train and needs much more data samples, which is a reason why 2D and 2.5D approximations were once preferred [31-33]. Here, the parameter explosion problem is alleviated by the U-Net structure and ResNet module (also named as 3D U-Net, V-Net) [37, 38].

As shown in Fig. 5, the FCN we utilized has a structure similar to the U-Net but with only two encode-decode stages. This is because the size of our network input was much smaller (48 * 48 * 48), and thus less encode-decode stages were applied to preserve spatial information. Note that each encoder or decoder of each stage contained two ResNet units rather than one to compensate for the less encoding-decoding stages. The proposed FCN has 20 cascading convolutional layers, and is comparably deep in the field of medical image analysis. A total of 12 skip connections was utilized, which contributes greatly to the neural network training.

It is known that FCNs were first introduced to resolve the object segmentation challenge. Although the challenge of object segmentation is deeply related to that of object detection, directly employing an FCN on detection problems would damage the detection capability. The largest difference between these two problems lies in the way the negative data is handled. For a typical segmentation scenario, such as volumetric medical images [39], the networks are trained with inputs that would contains positive labels most of the time. Inputs without positive labels are often neglected, which may deteriorate the network performance. In the following sections, three methods were proposed to try to address the mentioned challenges.

3.3.1 Multi-Step Training Scheme

In the proposed framework, inputs for FCNs were small cropped samples with a length l_{sample} rather than the entire volume due to the hardware limitations, as a typical 3D neural network is often excessively memory-consuming. It should be also emphasized that FCN does not have to deal with the entire abdomen, because the anatomic structures along the colon are similar. The detection complexity can be safely decreased as long as cropped inputs are big enough to contain the necessary contextual information for polyp detection.

Typical FCNs handle large-size inputs most of which covers all of the target area, and thus, they do not have to deal with the sampling strategy. However, inputs of our proposed FCN are smaller. A multistep training scheme was designed to sample the training data samples.

Warming up. Only positive data were sampled and fed into the FCN. The exclusion of negative data allowed the network to learn faster by concentrating on the polyp shape and context.

Strengthening. The network was trained with label-balance inputs. Negative data were randomly sampled from the non-polyp area in the cropped polyp pairs as in Fig. 3, which were relatively easier negative samples. We focused on cropped polyp pairs rather than the entire CT volumes. This is because the whole CT dataset (>500GBs) is too big to be entirely loaded into the computer memory, and on-the-fly complete CT volume loading can be time-consuming considering the many training epochs. Cropping from the cropped polyp pairs is quicker and the FCN is expected to retain detection capability under the assumption that polyps in the dataset are distributed in all the typical colon regions.

Refining. The half-trained FCN was applied to screen the whole volume pairs to generate predictions, and the generated FPs were integrated back into the training dataset, serving as high-quality negative data. During training, negative data were sampled from these generated hard FPs. Additionally, there is still a small number of negative samples sampled in the same way as that in strengthening, which is a regularization approach to maintain the training stability. Overall, the training data are label-balanced.

Repeating. The third step was repeated until the network converges to a stable state. The FCN is expected to learn more discriminative and representative features to differentiate the polyps from the hard FPs.

Unlike the boosting classification algorithm that allocates different weights to fixed data samples, training data were not fixed in our proposed training scheme. FPs were generated from the raw data every refining stage. It is expected that higher quality for the negative data selection should improve the detection capability for the proposed system.

3.3.2 Ensemble Loss

There are two important issues to be discussed in the selection of loss function for the polyp detection challenge.

One is the voxel label imbalance problem in positive labels. In detail, the imbalance lies between the number of voxels belonging to the polyp label and those belonging to the background area in a positive data sample. The former one in most cases is much smaller than the latter one.

The other issue is about the contradiction between the FP size and the misclassified cost. Generally, larger polyps are more diagnostically important, because they have a greater chance of being malignant tumors. Thus, bigger FPs should be suppressed less in order to avoid awful misses of polyp detections with a significant size. However, for typical FCN loss function for segmentation, the larger FPs often result in a higher misclassified cost for the FCN, which results in an easier rejection of larger predictions.

To resolve the first issue, as in many studies [38, 40], dice loss is introduced to overcome the imbalance of the voxels. Let $Y = \{y_i\}_{i=1}^{N}$ be the ground-truth polyp segmentation for a data sample consisting of N voxels, where $y_i = 1$ denotes that the ith voxel belongs to the polyps, while $y_i = 0$ is the background. Similarly, we assume that $Y = \{\hat{y}_i\}_{i=1}^{N}$ is the score map produced by the proposed FCN, where $\hat{y}_i = 1$ denotes that the ith voxel belongs to the polyps and vice versus. The dice loss is represented as

$$\mathcal{L}_{dice} = -\frac{2\sum_{i=1}^{N} y_i \hat{y}_i + \varepsilon_1}{\sum_{i=1}^{N} y_i^2 + \sum_{i=1}^{N} \hat{y}_i^2 + \varepsilon_2} \tag{1}$$

where ε_1 and ε_2 are small scalars to guarantee the numerical stability. When encountering a negative sample that does not have any voxels with positive labels, the dice loss collapses into an inverse function as:

$$\mathcal{L}_{collapsed\ dice} = -\frac{\varepsilon_1}{\sum_{i=1}^{N} \hat{y}_i^2 + \varepsilon_2} \tag{2}$$

It can be seen from the equations that the collapsed dice loss is unstable when the network has a small prediction on a small negative sample. The gradient of the collapsed loss decreases exponentially as the sum of the prediction increases, and consequently, larger FPs have considerably less contribution to the gradient of network training. That is, the bigger FPs are excluded from the training, hampering the FP detection performance of the larger polyps. A large ε_2 can be used to regularize the inverse function and to prevent the network from outputting smaller or even always empty results, but in this way, the normalizing ability of the dice loss for voxel imbalance is deteriorated.

A direct intuition to deal with the problem of collapsed dice loss is to utilize a hybrid loss with both dice loss and cross entropy loss as described previously [41]. However, it gives rise to the second issue discussed above. The convolutional filters share parameters and are updated together, which means that the larger FPs with more positive voxel predictions have a larger impact on the FCN training. Therefore, the trained FCN tends to eliminate the larger predictions, and the detection performance of the larger polyps suffers a dramatic decrease.

To resolve the controversy, a regularized cross entropy loss for the FPs is introduced. The regularized cross entropy loss function is as

$$\mathcal{L}_{cross\ entropy} = -\lambda\, \mathrm{H}\left(0.5 - \sum_{i=1}^{N} y_i\right) \frac{\sum_{i=1}^{N} (1 - y_i) \log(1 - \hat{y}_i)}{\sum_{i=1}^{N} \mathrm{H}(\hat{y}_i - 0.5) + \varepsilon_3} \tag{3}$$

where H is the Heaviside function, and ε_3 is a scalar to ensure the numerical stability. It should be emphasized that due to the Heaviside function, the denominator of the equation functions as a scalar in most cases, and $\mathcal{L}_{cross\_entropy}$ is only calculated for the negative samples. As a consequence of the loss

of regularization according to the size of prediction, the trained network is less likely to exclude the larger predictions because the cost of mistaking larger FPs as polyps is lower now.

For the positive samples, the dice loss with $\varepsilon_1 = 0$ is calculated. With the addition of the l2 regularization of the network parameters W, the proposed ensemble loss is summarized as follows:

$$\mathcal{L}_{ensemble\ loss} = \mathcal{L}_{dice} + \mathcal{L}_{cross\ entropy} + \beta\|W\|^2 \tag{4}$$

The proposed ensemble loss is capable of overcoming the imbalance of labels while respecting the nature of medical diagnostic analysis and, at the same time, avoiding numerical instability.

Generally speaking, an FCN system with dice loss function is mostly used for segmentation task. If directly used for detection task, it will inevitably bring performance decline. The proposed ensemble loss in this chapter can alleviate the problem, resulting in a more powerful CAD system.

3.3.3 Imbalanced Samples and Adaptive Data Transform in 3D Space

In addition to the label imbalance, there exists another type of data imbalance between the number of positive samples and that of the negative samples. In most cases, the number of negative samples can be tens of hundreds of times larger. Here, over-sampling of the positive samples is applied followed by the adaptive data transform as data augmentation to alleviate the imbalance problem.

As we know, the training of 3D CNNs requires a large number of data samples, a condition currently hard to meet in CTC. To generate as many data samples as possible, on-the-fly excessive free linear transform in the 3D space is applied to the input data before it is fed into the FCN. Scaling, rotation, flipping and translation are done in order.

To squeeze the last potential from the linear data augmentation, scaling and rotation, with continuous-value parameters, are introduced to produce as many as possible augmentations. Free rotation is allowed with random values in two degrees of freedom, while the ratio of the zooming σ has to comply with the following restriction:

$$\alpha < \sigma < \beta \tag{5}$$

$$\alpha = \max\left\{0.6, \min\left\{\frac{\rho_{min}}{\eta}, 0.9\right\}\right\} \tag{6}$$

$$\beta = \max\left\{1.1, \min\left\{\frac{\rho_{max}}{\eta}, 1.4\right\}\right\} \tag{7}$$

ρ_{max} denotes the diameter of the 10th largest polyp in the dataset, and ρ_{min} denotes the diameter of the 10th smallest polyp. Additionally, η is the diameter of the positive data processed. The zooming ratio is adaptively controlled to prevent the occurrence of unreasonably large or small polyps under the assumption that the identification of the polyps is scale-invariant in a certain range.

For the negative samples, the zooming ratio of fixed range is selected randomly.

4 Experiment

4.1 Implementation Details

The network was implemented in the deep learning framework of Tensorflow [42] utilizing three GPUs.

The l_{polyp_volume}, the length of the polyp volume pair, was set to 160. The l_{sample}, the length of the input data to the FCN, was set to 48 so that the close context of even the largest polyp in the dataset can be included. The length l_{test_window} was set to 160, and $l_{test\_overlap}$ was set to 32.

The learning rate was initialized as 0.0001, with an exponential decay of 0.95 every 1000 batches. The network was trained from scratch with weights initialized from the Gaussian distribution $\mathrm{N}(0, 0.05)$ and with offsets initialized as 0.001. During the test procedure, the exponential moving average of the parameters was applied to provide a more stable and robust prediction. For the multistep training scheme, there was only one refining stage applied, as we found that one refining stage was good enough in the case of CAD-CTC.

For the loss function, $\mathcal{E}_1$, $\mathcal{E}_2$, $\mathcal{E}_3$ was set to 0, 1e-4, 1e2, respectively. And λ was set to 1e4.

4.2 Evaluation Metrics

For the evaluation metrics, we employed the common practice of by-polyp sensitivity, per-scan sensitivity and FP rate. For a polyp in a CT scan, it was considered found as long as at least one voxel of the polyp mask is detected. The by-polyp sensitivity was the percentage of the successfully detected polyps among all the existing polyps, and a polyp was considered detected once it is found in either the supine or prone scan of the patient. The per-scan sensitivity was the percentage of the abnormal polyp-detected CT scans among all the CT scans, with the supine and prone scans considered as irrelative different scans. The FP rate was the number of detected FPs per CT scan. As most CAD-CTC systems are evaluated by the by-polyp sensitivity [15, 18, 32, 33], we abbreviated sensitivity for the by-polyp sensitivity in the following passages. A five-fold nested cross validation (CV) was applied so that every polyp was guaranteed to be tested.

4.3 Experiment Results

The Free-response Receiver Operating Characteristics (FROC) analysis of the proposed approach is shown in Fig. 8. Under an FP rate of 1.5, we achieved a sensitivity of **96.5%** for all the polyps, **99.1%** for large polyps (>=10mm), **95.3%** for small polyps (<10mm), **91.9%** for tiny polyps (<6mm), and a per-scan sensitivity of **90.5%**.Fig. 7 illustrates some successful detections, most of which are hard obstacles for classic CAD systems. It can be observed that the proposed method exhibited excellent performance on polyp detection under complex circumstances. Besides, the system also performed better on polyps obscured by stool and poor bowl preparation.

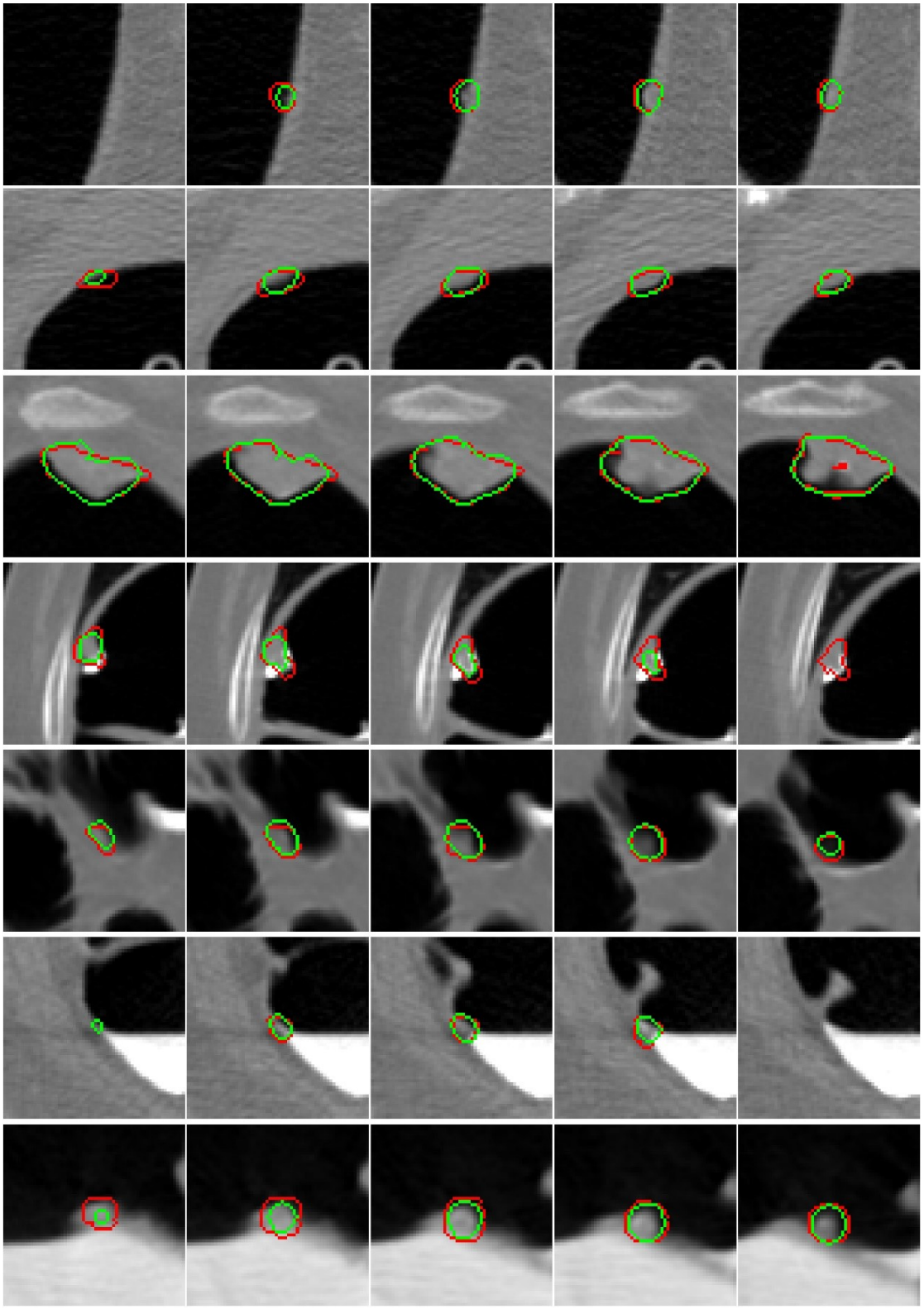

Fig. 7. Demonstration of seven successful detections. Each row shows a polyp in every other slice. The red curve is the outline of the ground truth, while the green line is the outline of the network detection. From top to bottom: (1) a small polyp with only a couple of voxels. (2) a flat polyp. (3) a large polyp with successful detection and segmentation. (4) a polyp half immersed in a drop of contrast fluid and on the colon fold. (5) a polyp under fuzzy and complex surroundings. (6) a polyp on the edge of the contrast fluid and air. (7) a polyp on the edge of a sharp colon fold, contrast fluid and air.

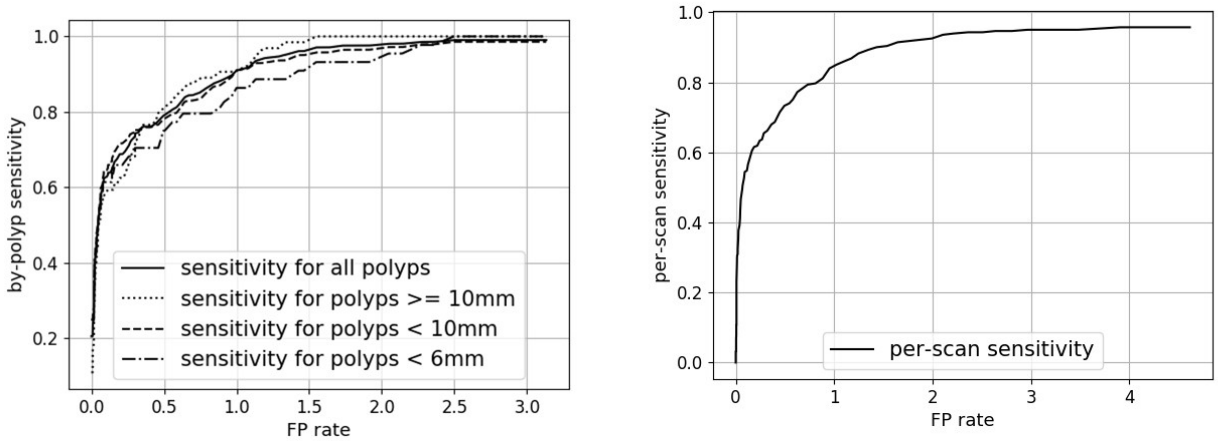

Fig. 8. FROC curves of the performance of the proposed framework. The left figure is the plot of the (by-polyp) sensitivity, while the right is the plot of per-scan sensitivity.

4.4 Comparison.

Several comparisons were carried out to demonstrate the efficacy of the proposed method, including two recently published CNN methods and a number of classic CAD frameworks.

4.4.1 CNN Methods.

As discussed in the introduction section, there are few CNN approaches for CTC in the public literature, which is partly due to the lack of available labeled data. Here, the comparison with two of the state-of-the-art CAD frameworks is listed below.

1) RVA

A similar 2.5D CNN scheme with an RVA strategy was implemented according to [33]. The method was evaluated on the same dataset in this paper. The comparison FROC curves are shown in Fig. 9. It can be observed that the 2.5D CNN with RVA was good at detecting large polyps (95.3% by-polyp sensitivity at 2.0 FPs per scan) but exhibited an inferior performance on smaller ones (87.6% by-polyp sensitivity at 2.0 FPs per scan) as reported in the original paper. Our proposed approach has an apparent advantage over small polyps.

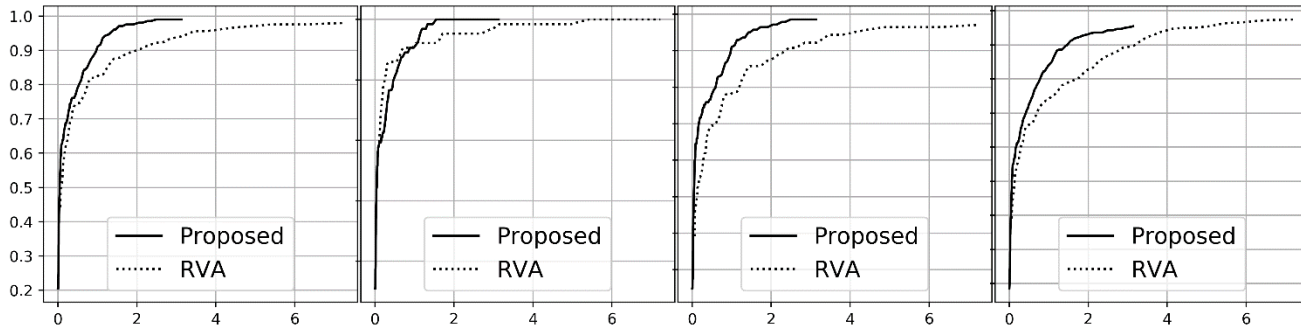

Fig. 9. FROC curves in comparison with RVA. The X axes represent the FP rate, and the Y axes represent the different sensitivities. The columns from left to right represent plots of different sensitivities, the sensitivities of all polyps, the sensitivities of polyps >= 10 mm, the sensitivities of polyps < 10 mm, and the per-scan sensitivities.

Table 3 Comparison with DEL

	By-polyp sensitivity at 3.9 FPs per scan (polyps >=10 mm)	By-polyp sensitivity at 2.9 FPs per scan (polyps<10 mm)	Number of polyps>=10 mm in the test set	Number of polyps<10 mm in the test set	Number of CTC patients
DEL	97.1%	86.8%	**69**	38	154
Proposed	**100.0%**	**98.6%**	64	**141**	**277**

Table 2 Comparison with a state-of-the-art CAD-CTC system

	sensitivity at 1.96 FP Rate	sensitivity at 1.1 FP Rate (>= 10 mm)	sensitivity at 2.7 FP Rate (< 10 mm)
Ren [33]	91.3%	93.7%	87.16%
Proposed	**96.4%**	**96.9%**	**98.6%**

2) Deep Transfer Learning (DEL) of Virtual Endoluminal Views

2D virtual endoluminal views rendered from colon segmentation were utilized to train the CNNs to reduce the number of FPs in [32]. The ensemble scheme relies on the output of a classic CAD-CTC system, which is difficult to reimplement. Therefore, a relative comparison was made, as shown in Table 2. For the detection of small polyps, our proposed method made use of much more polyp data and achieved a significantly better result while preserving a comparable performance for the detection of large polyps.

4.4.2 Classic Methods.

A dozen CAD-CTC systems have been presented for the last decade, most of which were evaluated on different proprietary datasets with different metrics. Considering the delicacy and complexity of typical CAD-CTC systems, it is impractical to reimplement all of them and make direct comparisons. To address the dilemma, two approaches were taken.

For one approach, a specific state-of-the-art CAD-CTC system, utilizing radiomic features, was selected and implemented [15], and a direct comparison under the same dataset in this paper was made in Table 3. The result indicates that the classic CAD system deteriorated when generalizing to more data, while ours still maintained a high performance. The advantage is believed to lie in the superior generalization capability of the CNN.

For the other approach, information of the CAD systems was gathered from the public literature to make a relative comparison. To keep the paper reasonably concise, only those methods meeting the following criteria were chosen: the number of polyps tested more than 50, published after 2005 and including polyps of <10 mm size. Additionally, papers published after 2010 were all included, except for those that were already compared above. It should be noted that Tulum [17] proposed a classic CAD-CTC system with an impressive high detection performance in 2017, but a relative comparison was inappropriate due to the fact that it was evaluated under common sensitivity and was validated on a small confidential dataset. Table 4 shows the comparisons between several collected methods. It can be concluded that the performance of the proposed framework was one of the best among all the existing systems, and that the proposed framework outperformed the systems with datasets of comparable sizes.

Table 4 List of recent CAD-CTC systems

Author	Year	Polyps for testing	Polyp sizes (mm)	Validation	By-polyp Sensitivity	FP rate
Näppi and Yoshida [43]	2007	40	>=6	Leave-one-patient-out	95%	3.6
		16	>=10	Leave-one-patient-out	94%	3.6
		44	>=6	Leave-one-patient-out	86%	4.2
		21	>=10	Leave-one-patient-out	100%	4.2
Yao, Li [44]	2009	43	>=10	Independent	93%	1.2
		112	6-9	Independent	76%	3.1
Li, Huang [45]	2009	121	>=6	Independent	77.4%	5.83
		30	>=10	Independent	91.3%	2.3
Oda, Kitasaka [46]	2009	57	>=6	Independent	91.2%	11.4
Van, van Ravesteijn [47]	2010	57	>=6	Leave-one-patient-out	95%	4
Wang, Yao [48]	2010	96	>=6	Leave-one-patient-out	83%	5
Mang, Bogoni [49-51]	2012	244	>=6	Independent	90.2%	2.33
Wang, Liang [18]	2015	130	5-8	Leave-one-patient-out	98%	2.2
Proposed	2018	64	>=10	Cross validation	99.1%	1.5
		141	3-9	Cross validation	95.3%	1.5
		205	all	Cross validation	96.5%	1.5

4.5 Module Analyses

The contribution of each key module in the proposed method was evaluated.

4.5.1 Role of the One-Stage Scheme

A two-step 3D CNN scheme was designed, including a 3D FCN for candidate screening and a 3D classification CNN for FP reduction. The screening network was implemented as [52]. The screening step achieved 99% sensitivity, with 21 FPs per volume. The CNN for the FP reduction step had a regular classification network structure, with several residual modules.

The experiment results are shown in the first row of Fig. 10. The two-step 3D CNN had a similar or even worse performance than the RVA scheme introduced above, and was apparently worse than the proposed one-stage scheme. It can be inferred that data inadequacy resulted in serious overfitting during the training of the 3D classification CNN from scratch, and that the design of the one-stage scheme alleviated the overfitting problem.

4.5.2 Role of the Multistep Training Strategy

The refining stage is the vital part to boost the performance of the FCN. To assess the effectiveness of the refining stage, system performances before and after the refining stage are shown in the second row of Fig. 10. It can be observed that the refining stage greatly enhanced the detection capability for the small polyps. The detection of smaller polyps is tougher and requires more attention of the network.

4.5.3 Role of the Ensemble Loss

Quantitative comparisons between dice loss, the unregularized ensemble loss, and the proposed ensemble loss, were made in the third row of Fig. 10. Dice loss with different settings for regularization

coefficients was employed to evaluate how the performance varied with the regularization coefficients. A plot of the ensemble loss without the regularization denominator was made to assess the contribution of the polyp size regularization.

It can be concluded from the graph that the proposed ensemble loss significantly improved the detection capability for big polyps as expected. Considering the higher diagnostic significance of larger polyps, the proposed ensemble loss reached a better balance between prioritizing the detection of big polyps and maintaining a high sensitivity for small ones. Moreover, the single dice loss was found instable when applied in the refining stage. The dice loss with a small ε_2 value tended to make the networks converge to an all-zero output.

4.5.4 Role of the Adaptive Data Transform

As is shown in the fourth row of Fig. 10, two ablation experiments are studied. They were applied with the same data transform as proposed but had a fixed range of scaling factor. The first one was scaled in a range of 0.6 - 1.5, which was the same as the largest range in the adaptive data transform, while the other one was scaled in a range of 0.9 - 1.1, which was the same as the smallest range. The experiments validated the necessity of adaptive scaling, as the performances of both of the other two settings deteriorated.

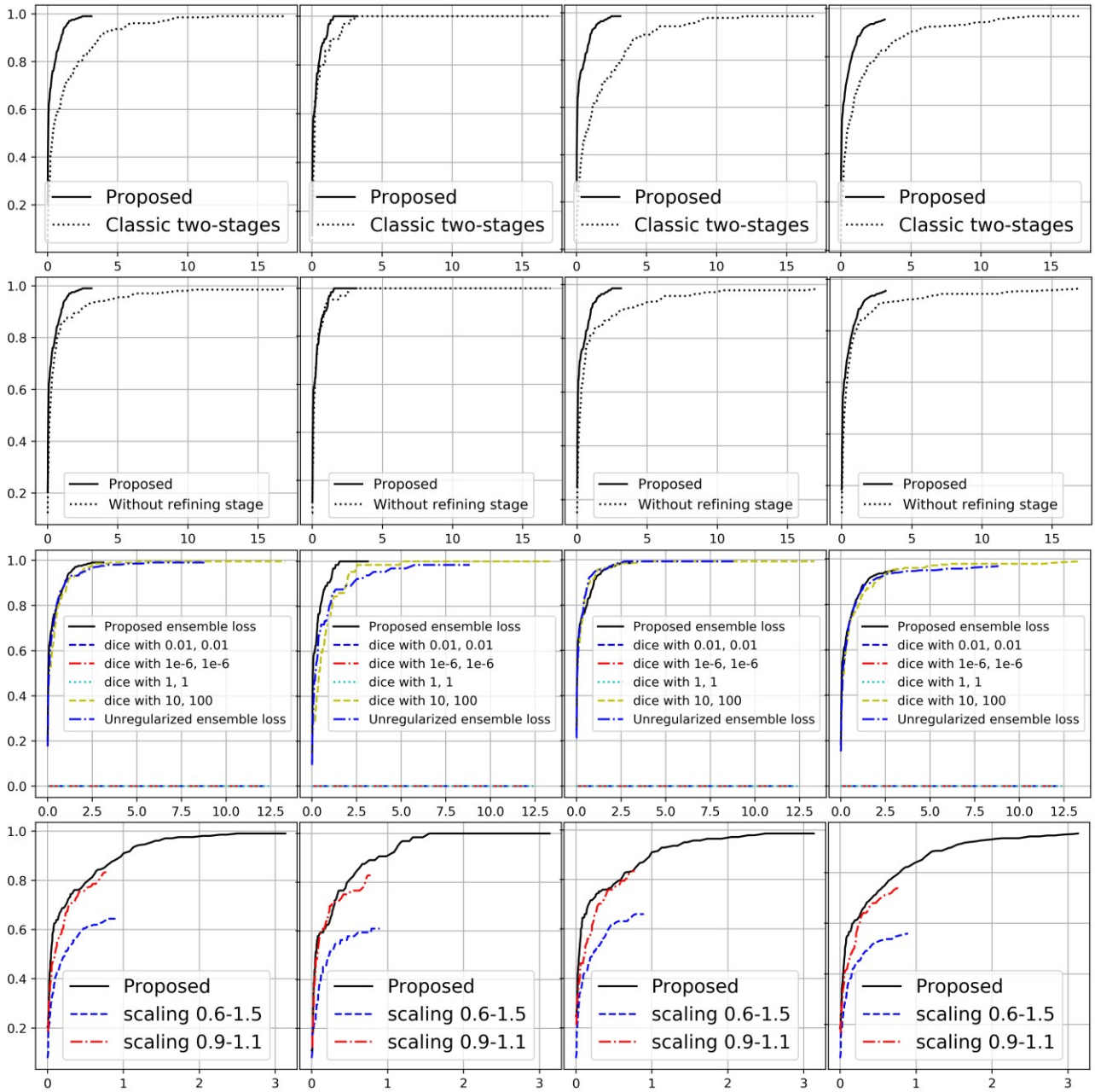

Fig. 10. FROC curves of sensitivities and FP rate in the comparison experiments. The X axes represent the FP rate, and the Y axes represent the different sensitivities. The columns from left to right represent plots of the different sensitivities, the sensitivities of all polyps, the sensitivities of polyps >= 10 mm, the sensitivities of polyps < 10 mm, and the per-scan sensitivities. The first row shows the results for section 4.5.1; the second one, for section 4.5.2; the third one, for 4.5.3; and the fourth one, for 4.5.4.

5. Discussion and Conclusion

The results of the experiment indicated that our proposed method achieved a superior performance even with a relatively simple end-to-end model, especially on small polyps. The RCNN series [24-26] is one of the most popular algorithms in the field of objection detection. The delicate region proposal and classification design of RCNNs is crucial for the complicated object detection of hundreds of types of objects. However, the polyp detection problem in CTC contains only two types of objects and can be well resolved by our proposed one-stage scheme.

In conclusion. we propose a high-performance one-stage CAD scheme for polyp detection in CTC. The simpler pipeline was made possible by the successful application and optimization of the 3D FCN detection module. The key to the optimization of FCNs for detection problems lies in the searching and integration of hard FPs during the network training. With the improvements to the data augmentation technique, data feeding scheme and loss function, the detection capability of the network is boosted to find the more discriminative geometrical and semantic features in the natural 3D space. Thorough comparisons and extensive experiments indicated that our adaptation and optimization of the FCNs for the detection scenario indeed improved the performance of the system, and our one-stage CAD framework was effective at detecting polyps of various shapes under different circumstances.

Ethical Statement and Acknowledgements

This work was supported by National Natural Science Foundation of China (No. 81371634), National Key Research and Development Program (No. 2016YFC0104608), National Basic Research Program of China (No. 2010CB834302), and Shanghai Jiao Tong University Medical Engineering Cross Research Funds (No. YG2017ZD10, YG2014ZD05). The authors declare that they have no conflict of interest.

This article does not contain any studies with human participants performed by any of the authors.

Reference

1. Society AC. Cancer Facts & Figures 2018. American Cancer Society Atlanta, GA; 2018.
2. Johnson CD, Dachman AH. CT colonography: The next colon screening examination? Radiology. 2000;216(2):331-41.
3. Näppi JJ, Frimmel H, Dachman AH, Yoshida H. Computerized detection of colorectal masses in CT colonography based on fuzzy merging and wall‐thickening analysis. Medical physics. 2004;31(4):860-72.
4. Summers RM, Beaulieu CF, Pusanik LM, Malley JD, Jeffrey Jr RB, Glazer DI, Napel S. Automated polyp detector for CT colonography: feasibility study. Radiology. 2000;216(1):284-90.
5. Kiss G, Van Cleynenbreugel J, Thomeer M, Suetens P, Marchal G. Computer-aided diagnosis in virtual colonography via combination of surface normal and sphere fitting methods. European Radiology. 2002;12(1):77-81.
6. Yoshida H, Nappi J. Three-dimensional computer-aided diagnosis scheme for detection of colonic polyps. IEEE transactions on medical imaging. 2001;20(12):1261-74.
7. Paik DS, Beaulieu CF, Rubin GD, Acar B, Jeffrey RB, Yee J, Dey J, Napel S. Surface normal overlap: a computer-aided detection algorithm with application to colonic polyps and lung nodules in helical CT. IEEE transactions on medical imaging. 2004;23(6):661-75.
8. Näppi J, Dachman AH, MacEneaney P, Yoshida H. Automated knowledge-guided segmentation of colonic walls for computerized detection of polyps in CT colonography. Journal of computer assisted tomography. 2002;26(4):493-504.
9. Konukoglu E, Acar B, Paik DS, Beaulieu CF, Rosenberg J, Napel S. Polyp enhancing level set evolution of colon wall: method and pilot study. IEEE Transactions on Medical Imaging. 2007;26(12):1649-56.
10. Li P, Napel S, Acar B, Paik DS, Jeffrey Jr RB, Beaulieu CF. Registration of central paths and colonic polyps between supine and prone scans in computed tomography colonography: Pilot study: Registration of colonic data in CT colonography. Medical physics. 2004;31(10):2912-23.
11. Näppi J, Okamura A, Frimmel H, Dachman A, Yoshida H. Region-based Supine-prone Correspondence for the Reduction of False-positive CAD Polyp Candidates in CT Colonography1. Academic Radiology. 2005;12(6):695-707.
12. Acar B, Beaulieu CF, Gokturk S, Tomasi C, Paik DS, Jeffrey B, Yee J, Napel S. Edge displacement field-based classification for improved detection of polyps in CT colonography. IEEE Transactions on Medical Imaging. 2002;21(12):1461-7.
13. Gokturk SB, Tomasi C, Acar B, Beaulieu CF, Paik DS, Jeffrey RJ, Yee J, Napel S. A statistical 3-D pattern processing method for computer-aided detection of polyps in CT colonography. IEEE Transactions on Medical Imaging. 2001;20(12):1251-60.
14. Van Wijk C, Van Ravesteijn VF, Vos FM, Van Vliet LJ. Detection and segmentation of colonic polyps on implicit isosurfaces by second principal curvature flow. IEEE Transactions on Medical Imaging. 2010;29(3):688-98.
15. Ren Y, Ma J, Xiong J, Lu L, Zhao J. High-Performance CAD-CTC Scheme Using Shape Index, Multiscale Enhancement Filters, and Radiomic Features. IEEE Transactions on Biomedical Engineering. 2017;64(8):1924-34.
16. Ren Y, Ma J, Xiong J, Chen Y, Lu L, Zhao J. Improved False Positive Reduction by Novel Morphological Features for Computer-Aided Polyp Detection in CT Colonography. IEEE Journal of Biomedical & Health Informatics. 2018;PP(99):1-.

17. Tulum G, Bolat B, Osman O. A CAD of fully automated colonic polyp detection for contrasted and non-contrasted CT scans. International journal of computer assisted radiology and surgery. 2017;12(4):627-44.
18. Wang H, Liang Z, Li LC, Han H, Song B, Pickhardt PJ, Barish MA, Lascarides CE. An adaptive paradigm for computer-aided detection of colonic polyps. Physics in Medicine & Biology. 2015;60(18):7207.
19. Wang H, Song W, Li L, Cao Y, Pan H, Ma M, Huang J, Mao G, Liang Z, editors. A Novel Approach on the Colon Wall Segmentation and Its' Application. International Workshop on Computer-Assisted and Robotic Endoscopy; 2014: Springer.
20. Lu L, Zhang D, Li L, Zhao J. Fully automated colon segmentation for the computation of complete colon centerline in virtual colonoscopy. IEEE Transactions on Biomedical Engineering. 2012;59(4):996-1004.
21. Cai W, Lee J-G, Zhang D, Kim SH, Zalis M, Yoshida H. Electronic cleansing in fecal-tagging dual-energy CT colonography based on material decomposition and virtual colon tagging. IEEE Transactions on Biomedical Engineering. 2015;62(2):754-65.
22. Yao J, Summers RM. Adaptive deformable model for colonic polyp segmentation and measurement on CT colonography. Medical physics. 2007;34(5):1655-64.
23. Krizhevsky A, Sutskever I, Hinton GE, editors. Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems; 2012.
24. Girshick R, Donahue J, Darrell T, Malik J, editors. Rich feature hierarchies for accurate object detection and semantic segmentation. Proceedings of the IEEE conference on computer vision and pattern recognition; 2014.
25. Girshick R. Fast R-CNN. Computer Science. 2015.
26. Ren S, He K, Girshick R, Sun J, editors. Faster R-CNN: towards real-time object detection with region proposal networks. International Conference on Neural Information Processing Systems; 2015.
27. Dou Q, Chen H, Yu L, Qin J, Heng P-A. Multilevel contextual 3-d cnns for false positive reduction in pulmonary nodule detection. IEEE Transactions on Biomedical Engineering. 2017;64(7):1558-67.
28. Xiang L, Wang Q, Nie D, Zhang L, Jin X, Qiao Y, Shen D. Deep embedding convolutional neural network for synthesizing CT image from T1-Weighted MR image. Medical image analysis. 2018;47:31-44.
29. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, van der Laak JA, Van Ginneken B, Sánchez CI. A survey on deep learning in medical image analysis. Medical image analysis. 2017;42:60-88.
30. Tachibana R, Näppi JJ, Hironakaa T, Kim SH, Yoshida H, editors. Deep learning for electronic cleansing in dual-energy CT colonography. Medical Imaging 2016: Computer-Aided Diagnosis; 2016: International Society for Optics and Photonics.
31. Näppi JJ, Hironaka T, Regge D, Yoshida H, editors. Deep transfer learning of virtual endoluminal views for the detection of polyps in CT colonography. Medical Imaging 2016: Computer-Aided Diagnosis; 2016: International Society for Optics and Photonics.
32. Umehara K, Näppi JJ, Hironaka T, Regge D, Ishida T, Yoshida H, editors. Deep ensemble learning of virtual endoluminal views for polyp detection in CT colonography. Medical Imaging 2017: Computer-Aided Diagnosis; 2017: International Society for Optics and Photonics.
33. Roth HR, Lu L, Liu J, Yao J, Seff A, Cherry K, Kim L, Summers RM. Improving computer-aided detection using convolutional neural networks and random view aggregation. IEEE transactions on medical imaging. 2016;35(5):1170-81.
34. Smith K, Clark K, Bennett W, Nolan T, Kirby J, Wolfsberger M, Moulton J, Vendt B, Freymann J. Data From CT_COLONOGRAPHY. 2015.

35. Johnson CD, Chen MH, Toledano AY, Heiken JP, Dachman A, Kuo MD, Menias CO, Siewert B, Cheema JI, Obregon RG. Accuracy of CT colonography for detection of large adenomas and cancers. New England Journal of Medicine. 2008;359(12):1207.
36. Clark K, Vendt B, Smith K, Freymann J, Kirby J, Koppel P, Moore S, Phillips S, Maffitt D, Pringle M. The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository. Journal of Digital Imaging. 2013;26(6):1045-57.
37. Ronneberger O, Fischer P, Brox T, editors. U-net: Convolutional networks for biomedical image segmentation. International Conference on Medical image computing and computer-assisted intervention; 2015: Springer.
38. Milletari F, Navab N, Ahmadi S-A, editors. V-net: Fully convolutional neural networks for volumetric medical image segmentation. 3D Vision (3DV), 2016 Fourth International Conference on; 2016: IEEE.
39. Dou Q, Yu L, Chen H, Jin Y, Yang X, Qin J, Heng P-A. 3D deeply supervised network for automated segmentation of volumetric medical images. Medical image analysis. 2017;41:40-54.
40. Lian C, Zhang J, Liu M, Zong X, Hung SC, Lin W, Shen D. Multi-Channel Multi-Scale Fully Convolutional Network for 3D Perivascular Spaces Segmentation in 7T MR Images. Medical Image Analysis. 2018;46:106.
41. Qi D, Cheng O, Cheng C, Hao C, Heng PA. Unsupervised Cross-Modality Domain Adaptation of ConvNets for Biomedical Image Segmentations with Adversarial Loss. 2018.
42. Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J, Devin M, Ghemawat S, Irving G, Isard M, editors. Tensorflow: a system for large-scale machine learning. OSDI; 2016.
43. Näppi J, Yoshida H. Fully automated three-dimensional detection of polyps in fecal-tagging CT colonography. Academic radiology. 2007;14(3):287-300.
44. Yao J, Li J, Summers RM. Employing topographical height map in colonic polyp measurement and false positive reduction. Pattern Recognition. 2009;42(6):1029-40.
45. Li J, Huang A, Yao J, Liu J, Van Uitert RL, Petrick N, Summers RM. Optimizing computer‐aided colonic polyp detection for CT colonography by evolving the Pareto front a. Medical physics. 2009;36(1):201-12.
46. Oda M, Kitasaka T, Mori K, Suenaga Y, Takayama T, Takabatake H, Mori M, Natori H, Nawano S. Digital bowel cleansing free colonic polyp detection method for fecal tagging CT colonography. Academic radiology. 2009;16(4):486-94.
47. Van WC, van Ravesteijn VF, Vos FM, van Vliet LJ. Detection and segmentation of colonic polyps on implicit isosurfaces by second principal curvature flow. IEEE Transactions on Medical Imaging. 2010;29(3):688-98.
48. Wang S, Yao J, Petrick N, Summers RM. Combining Statistical and Geometric Features for Colonic Polyp Detection in CTC Based on Multiple Kernel Learning. International Journal of Computational Intelligence & Applications. 2010;9(1):1.
49. Mang T, Bogoni L, Salganicoff M, Wolf M, Raykar V, Macari M, Pickhardt PJ, Iafrate F, Laghi A, Weber M. Computer-aided detection of colorectal polyps in CT colonography with and without fecal tagging: a stand-alone evaluation. Investigative Radiology. 2012;47(2):99-108.
50. Näppi J, Yoshida H. Feature‐guided analysis for reduction of false positives in CAD of polyps for computed tomographic colonography. Medical Physics. 2003;30(7):1592-601.
51. Summers RM, Johnson CD, Pusanik LM, Malley JD, Youssef AM, Reed JE. Automated polyp detection at CT colonography: feasibility assessment in a human population. Radiology. 2001;219(1):51-9.
52. Chen Y, Ren Y, Fu L, Xiong J, Larsson R, Xu X, Sun J, Zhao J, editors. A 3D Convolutional Neural Network Framework for Polyp Candidates Detection on the Limited Dataset of CT

Colonography. 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); 2018: IEEE.