IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

SPECIAL SECTION ON DEEP LEARNING FOR COMPUTER-AIDED MEDICAL DIAGNOSIS

Received January 31, 2019, accepted February 13, 2019, date of publication February 21, 2019, date of current version March 12, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2900672

Ensemble of Instance Segmentation Models for Polyp Segmentation in Colonoscopy Images

JAEYONG KANG[1] AND JEONGHWAN GWAK[1,2,3]

[1]Biomedical Research Institute, Seoul National University Hospital, Seoul 03080, South Korea
[2]Department of Radiology, Seoul National University Hospital, Seoul 03080, South Korea
[3]Department of Radiology, College of Medicine, Seoul National University, Seoul 03080, South Korea

Corresponding author: Jeonghwan Gwak (james.han.gwak@gmail.com)

This work was supported in part by the Basic Science Research Program through the National Research Foundation of Korea (NRF), the Ministry of Education, under Grant NRF-2017R1D1A1B03036423, in part by the Brain Research Program through the NRF, the Ministry of Science, ICT and Future Planning (MSIT), under Grant NRF-2016M3C7A1905477 and Grant NRF-2019M3C7A1020406, and in part by the Engineering Research Center (ERC) Program of Extreme Exploitation of Dark Data through the Korean Government (MSIT) under Grant NRF-2018R1A5A1060031.

ABSTRACT Colorectal cancer is the second most frequently diagnosed cancer in women and the third most frequently diagnosed cancer in men. At least 80%–95% of the colorectal cancers are evolved from intestinal polyps. Although colonoscopy is regarded as the most effective method for screening and diagnosis, the success of the procedure is highly dependent on the level of hand-eye coordination and the operator skills. Thus, we are primarily motivated by the need for obtaining an early and accurate diagnosis of polyps in the colonoscopy images. In this paper, we employed the powerful object detection neural network ''Mask R-CNN'' to identify and segment polyps in the colonoscopy images. Also, we proposed an ensemble method to combine the two Mask R-CNN models with different backbone structures (ResNet50 and ResNet101) to enhance the performance. Mask R-CNNs in our model were first trained on COCO dataset, and then finely tuned using intestinal polyp dataset since a large number of annotated colonoscopy images are not easily accessible. In order to evaluate our proposed model, we used three open intestinal polyp datasets, CVC-ClinicDB, ETIS-Larib, and CVC-ColonDB. Our results show that our transfer learning-based ensemble model significantly outperforms state-of-the-art methods.

INDEX TERMS Polyp segmentation, transfer learning, medical image analysis, deep learning, machine learning, artificial intelligence.

I. INTRODUCTION

Colorectal cancer is the third most frequently diagnosed cancer after lung cancer and breast cancer [1]. Therefore, prevention of colorectal cancer by detecting and removing colorectal adenomas (i.e., paraneoplastic lesions) is of paramount importance and has become a global public health priority. The prevention of colorectal cancer has been mostly done with the help of regular colonoscopy screenings. Colonoscopy is an endoscopic procedure for the screening and diagnosis of colorectal cancer. During this procedure, a long, flexible tube that has a light and a camera on the end is inserted into the rectum to remove polyps or other types of abnormal tissue inside of the entire colon. However, depending on the type and size, roughly 8-37% of polyps are missed during colonoscopic examination [2]. Missed polyps are potential precursors to colorectal cancer, which causes the third most commonly occurring cancer globally [3]. Thus, we conducted the recent research to develop an automatic polyp segmentation system that can encourage clinical endoscopists to detect tiny and flat polyps more effectively.

Several research groups have developed computer-aided system for automatic polyp segmentation to provide early clues of colorectal cancers. Some of these systems are focused on reducing the inspection time and segmenting polyp region using algorithms such as fuzzy clustering, region growing, level-set, and machine learning techniques (e.g., linear and quadratic discriminant classifiers and neural network classifiers) [4]. Polyp segmentation is challenging given the polyps texture and shape. Some computer-aided polyp segmentation systems use texture and morphological features to differentiate the polyp regions from folding

The associate editor coordinating the review of this manuscript and approving it for publication was Yudong Zhang.

regions on the colon wall which mimic them [5]. More recently, deep learning models such as convolutional neural networks have been applied to the task of automatic poly segmentation with promising results [6], [7].

Mask R-CNN [8] is one of the best deep-learning models for instance segmentation which first detects targets in the image, and produce the predicted mask for each detected target. Mask R-CNN outperforms other approaches in some of the COCO's challenging tasks, including person keypoint detection, instance segmentation, and bounding-box object detection. However, defining and training a Mask R-CNN architecture from the beginning is not a trivial task. The training process typically requires an excessive amount of labeled data, something that is generally difficult to be obtained for colonoscopy images. To overcome this problem, a transfer learning technique can be used. Transfer learning is a machine learning technique where information that is learned in one setting is exploited to improve generalization in another setting.

In this work, we proposed a transfer learning-based ensemble model for polyp segmentation. In our model, we ensembled the two Mask R-CNN models with different backbone structures including ResNet50 and ResNet101 to get better performance. In order to achieve polyp segmentation on colonoscopy images, Mask R-CNNs in our model were first trained on COCO dataset, and then fine-tuned using intestinal polyp dataset. We validated our method using well-known three public available datasets: CVC-ClinicDB [27] and ETIS-Larib [10] from the MICCAI 2015 polyp detection challenge [9], and CVC-ColonDB [28]. To this end, the contributions of this work are as follows:

1) We presented transfer learning based Mask R-CNN for polyp segmentation. To the best of our knowledge, this is the first work to use Mask R-CNN for the task of polyp segmentation.
2) We presented an ensemble method to combine the two Mask R-CNN models with different backbone structures (ResNet50 and ResNet101) to get better performance.
3) We demonstrate that our proposed method outperform state-of-the-art methods using dataset from MICCAI 2015 polyp detection challenge.

The rest of this work is organized as follows. In Section 2, we briefly present related works on polyp segmentation approaches. Then, we describe our proposed method for polyp segmentation in Section 3. The experimental results are presented in Section 4. Finally, Section 5 summarizes and concludes this work.

II. RELATED WORK

The first approach for polyp segmentation is using image processing methods. Many image processing methods have been applied to various medical domain such as brain imaging [41]–[44] and polyp segmentation [10]–[13], [45], [46]. Karargyris and Bourbakis [45] extracted features by performing Log–Gabor filters for automatic polyp segmentation. Jia [46] used K-means clustering method to localize and segment polyp contours. Bernal *et al.* [10] proposed method using depth of valleys (DoV) image for automatic polyp segmentation. They used the watershed algorithm to segment images into polyp candidate regions and then classifies each region into polyp and non-polyp. This classification is based on regions information and DoV in each region. Region information contains mean and standard deviation of each region and DoV is based on calculation of eigenvalues and eigenvectors of the gradient image. Ganz *et al.* [11] proposed a method based on Hough transform to find region of interest (ROI) and specular reflection suppression with exemplar-based image inpainting as a preprocessing method. After preprocessing, they use a method using ultrametric contour map (UCM), called shape-UCM [12] for image segmentation. Shape-UCM works based on image gradient contours and spectral clustering. After performing shape-UCM algorithm, they use a scheme to improve edges resulted from the shape-UCM algorithm. Also, they use images transformed into LAB color space and the image texture as a feature to refine edges. Ellipse fitting method is used to extract polyp regions from all candidate regions. The method proposed in [13] uses an improved watershed algorithm, called ''marker-controlled watershed'' method, as the initial stage for segmenting polyps. Hwang *et al.* [13] also use the region-maxima method for selection of an initial point in the watershed algorithm. After that, they use elliptical fitting to discard unwanted regions resulted in the previous step.

The second approach in polyp segmentation is feature extraction from image patches and labeling of patches as polyp and non-polyp based on extracted features. Tajbakhsh *et al.* [14] suggested a method based on Canny edge detector in each of the three color channels of the image and the work of [15]. This is done to construct edge maps. After that, oriented patches for each pixel are extracted to classify them as polyp or non-polyp. The feature extraction method proposed in [14] extracts sub-patches with 50% overlap and calculates their average vertically, which results in 1D intensity signal. After that, they used discrete cosine transform (DCT) coefficients as a feature for each extracted patch. Finally, they used the two-stage random forest classifier to assign a label to each patch. The first stage of the classifier converts low-level features into high-level features and feed them to the second stage classifier for classifying each patch into polyp and non-polyp.

The third approach for polyp segmentation is using Convolutional Neutral Networks (CNN). CNN is a well-known deep learning architecture where complex features of raw images are extracted by applying trainable filters and pooling operations [16]. In CNN, the extracted features are fed into a subsequent classifier. CNN has been successfully applied to various object detection tasks [30]–[32], classification tasks [33]–[35], and medical image segmentation task such as coronary artery calcifications [36], the pancreas [37], brain regions [38], [39], knee cartilage [40], and polyp regions [17]–[19]. In [17], they analyzed CNN results to

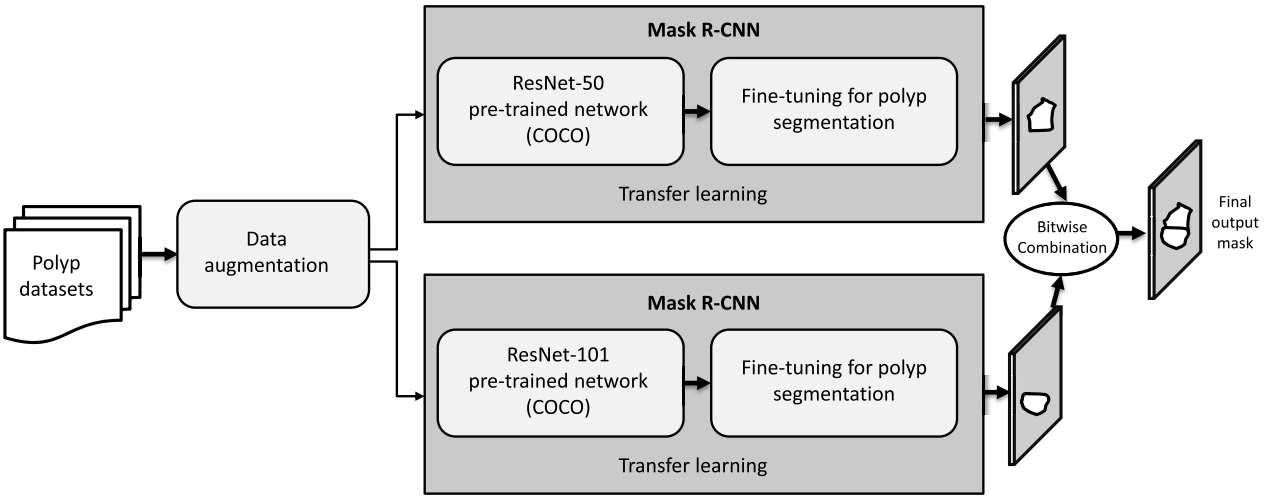

FIGURE 1. Overall architecture of our proposed framework.

see whether a fine-tuned CNN, or CNN from scratch, performs better in medical imaging applications such as polyp detection in colonoscopy videos. They showed that fine-tuned CNN works better than CNN from scratch. Park *et al.* [7] uses CNN as a feature extractor in three scales patch representation to segment polyp region. CNN calculates 60 features for each input patch, and then uses fully-connected layer with 256 neurons for classification of each input patch. Moreover, Gaussian filter is employed to smooth the segmentation results and decrease noise after performing CNN. Ribeiro *et al.* [19] uses two pooling layers and three convolution layers to extract features from RGB patches, and fully-connected layer to classify 1024 extracted features.

The fourth approach for polyp segmentation is using Fully Convolutional Networks (FCN) [20]. Over the last few years, FCN is one of the best deep learning algorithms for improving polyp segmentation, because of their computational efficiency for dense prediction. FCN is new generation of CNNs, which is implemented by replacing fully connected layer with deconvolution and using the information of previous layers for increasing segmentation accuracy. Chen *et al*. [47] combined the fully convolutional neural network with a fully connected conditional random field to sharpen object boundaries and improve segmentation performance. Chen *et al*. [18] then improved their network by using pooling operations and convolution filters at multiple rates and multiple effective fields-of-view, to extract better multi-scale contextual information. After that, they used the second network to refine the segmentation regions, gradually recover the spatial information, and sharpen object boundaries. In the field of polyp segmentation, Akbari *et al.* [21] adopted FCN to find potential polyp candidates, and then segment polyp regions by using the Otsu thresholding method which significantly improves segmentation accuracy. The method proposed in [22] use FCN for segmentation of polyp region candidates. Those candidates are further refined by using texton features and random forest classifier. Texton features are obtained by using k-means clustering on the convolution of input patch and Gabor filter bank for different orientations.

In this study, we compare our proposed method for polyp segmentation with FCN with six different architectures: AlexNet [23], GoogLeNet [24], VGG [25] and three versions of the ResNets architecture with 50, 101 and 152 layers of depth [26].

III. METHODS

In this section, we first describe the architecture of the proposed method. Then we outline the details of three key components in following subsections.

The overview of the proposed method can be seen in Figure 1. The proposed method consists of three components: 1) data augmentation, 2) two Mask R-CNNs with different backbone structures (Resnet50 and 101) pre-trained on the COCO dataset, and 3) the bitwise combination of two masks to enhance the segmentation performance as our ensemble method.

A. DATA AUGMENTATION

For an automated polyp segmentation method using deep learning networks, data augmentation is an essential step in improving segmentation performance. Because the endoscopy procedures involving moving camera control, color calibration are not consistent, the appearance of endoscopy images significantly changes across different laboratories. The data augmentation step brings endoscopy images into an extended space that can cover all their variances. Since access data is limited due to privacy concerns, the deep networks used for polyp segmentation were often trained with insufficient training datasets. As a consequence, the polyp segmentation performance is hindered by this lack of training data. Recent work has demonstrated the effectiveness of data augmentation in increasing the

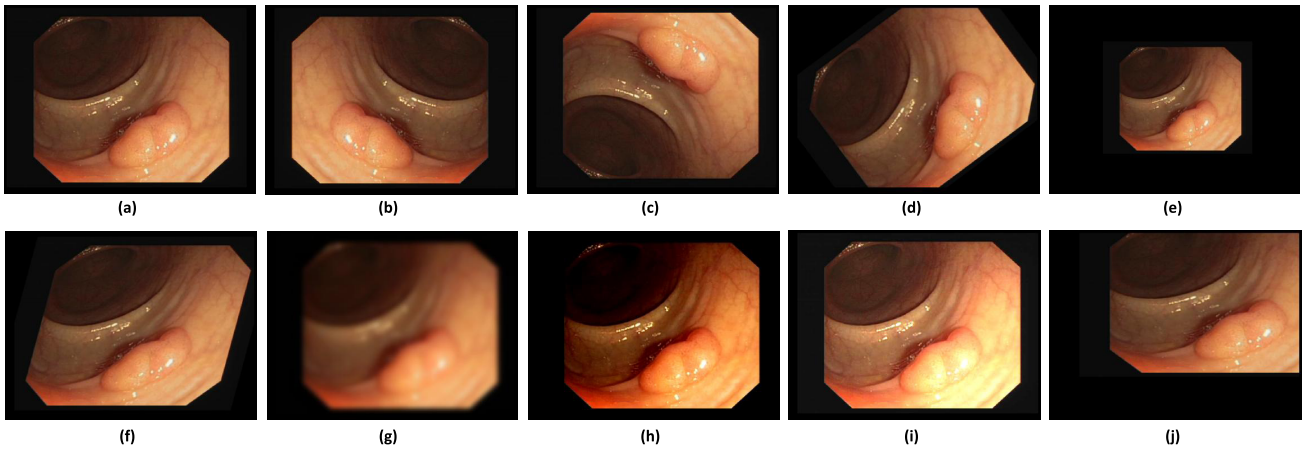

FIGURE 2. Examples of data augmentation used in this work.

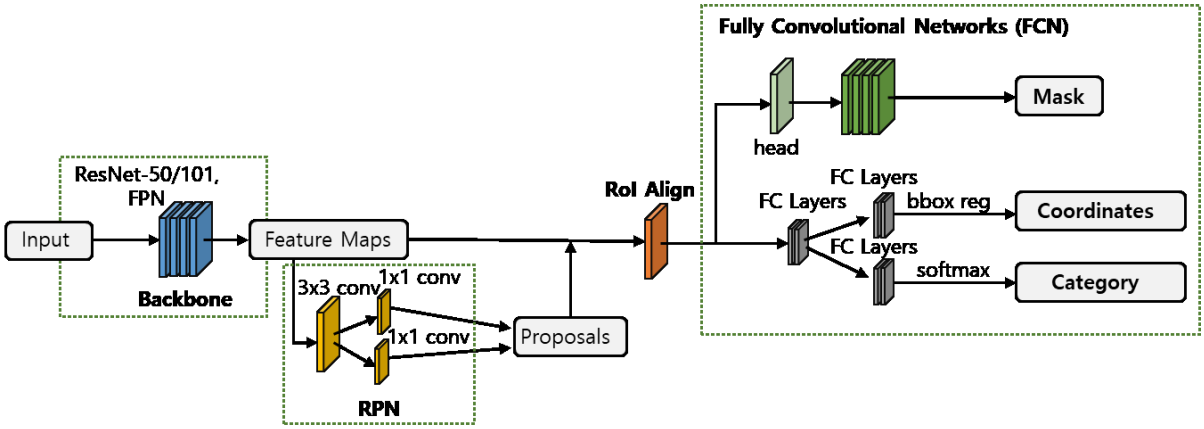

FIGURE 3. Detailed framework of Mask R-CNN.

amount of training data based on our original limited training dataset. By augmenting training data, we can also reduce the over-fitting problem on training models. Figure 2 shows the examples of data augmentation method applied to the original polyp image (Figure 2.a). In our model, the applied methods of augmentation are: Vertical flipping (Figure 2.b), horizontal flipping (Figure 2.c), random rotation between -45 and 45 degrees (Figure 2.d), random scaling ranging from 0.5 to 1.5 (Figure 2.e), random shearing between -16 and 16 degrees (Figure 2.f), random Gaussian blurring with a sigma of 3.0 (Figure 2.g), random contrast normalization by a factor of 0.5 to 1.5 (Figure 2.h), random brightness ranging from 0.8 to 1.5 (Figure 2.i), and random cropping and padding by 0–25% of height and width (Figure 2.j).

B. MASK R-CNN WITH TRANSFER LEARNING

Mask R-CNN [8] is a flexible, small generic object instance segmentation method which first detects targets in the image, and produces a high-quality segmentation result for each target. It is an extension to the Faster R-CNN [29], and adds a new branch to predict an object mask that is parallel with bounding box recognition branch. Also, it is easily extended to other tasks, such as person keypoint recognition for estimating a person's posture. It has achieved the best results in some of the COCO's challenging tasks, including person keypoint detection, bounding-box object detection, and instance segmentation.

In our proposed framework, we used two different backbone structures (ResNet50 and ResNet101) for each Mask R-CNN. The detailed framework of Mask R-CNN was illustrated in Figure 3. Mask R-CNN is a two-stage framework. The first stage is region proposal networks (RPN). RPN is a new proposal generation network from Faster R-CNN. It replaces the selective search algorithm in the Fast R-CNN and previous R-CNN. Also, it integrates all the content in one network to improve the detection speed. The second stage has two parallel branches. The first one is the bounding box branch for detection. It contains bounding box regression and classification. The second one is the mask branch for segmentation. In Mask R-CNN, the loss function is defined as follows:

$$L = L_{cls} + L_{box} + L_{mask}, \tag{1}$$

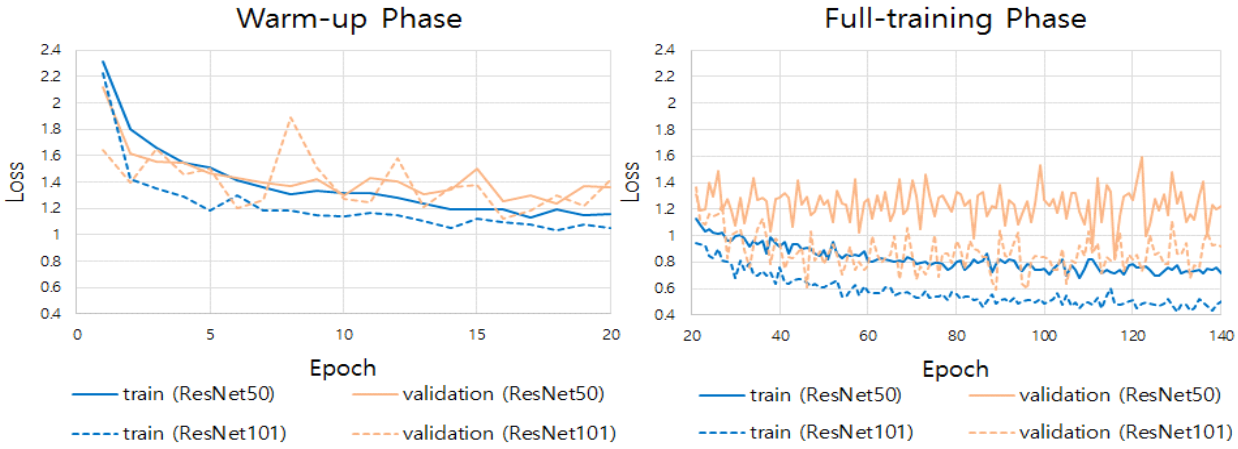

FIGURE 4. The two loss graphs for the warm-up phase and the full-training phase.

where L is loss, L_{cls} is the classification loss, L_{box} is the box regression loss, and L_{mask} is the mask loss. Loss for classification and box regression is same as Faster R-CNN [29]. To each mask, a per-pixel sigmoid is applied. The mask loss is then defined as the average binary cross entropy loss.

Despite the clear advantages that Mask R-CNN offers, defining and training a Mask R-CNN architecture from the beginning is not a trivial task. The training process typically requires an excessive amount of labeled data, something that is generally difficult to be obtained for colonoscopy images. In addition, selecting the architecture that provides the best compromise between inference ability and the convergence speed is both challenging and time consuming. Finally, potential issues of convergence and overfitting often require significant tuning of the learning parameters [17]. Alternatively, it has been suggested [17] that fine-tuning deep-learned architectures for specific tasks (e.g. image segmentation and detection), provides better performance than designing and training a new network model. Such an approach can be applied in segmentation tasks involving colonoscopy images, despite being substantially different from the natural images used in training the initial network, by using a smaller set of images to retrain (fine-tune) a pre-existing network. We adhere to this strategy using COCO dataset since a large number of annotated colonoscopy images are not easily accessible.

C. ENSEMBLE METHOD

Deep neural networks are notorious for having extremely high-dimensional, non-convex loss functions with many local minima. If Mask R-CNN was initialized with different pre-trained model with different backbone structure, the network is therefore virtually guaranteed to converge to different solutions, although it uses the same training data. For instance, Mask R-CNN with ResNet101 produced better segmentation results than Mask R-CNN with ResNet101 for some polyp images, and vice versa. Based on this observation, we use an ensemble method that combines two predicted masks by bitwise operation as illustrated in Figure 1.

D. TRAINING DETAILS

The training set is split into 90 % for learning the weights and 10 % for validating our model during the training step. Training of our proposed model is divided into two phases: (1) the warm-up phase and (2) the full-training phase. Furthermore, all compared algorithms have been programmed/trained using Keras and Tensorflow backend on a PC with a GTX-Titan X GPU. Also, the two loss graphs are shown in Figure 4.

1) Warm-up Phase: Pre-trained weights of the backbone model (ResNet50 or ResNet101) using COCO dataset are loaded as the training begins. Furthermore, we temporarily freeze the backbone model and update only the rest part of the network via stochastic gradient descent (SGD) with momentum. The learning rate, weight decay, and momentum of SGD is set to 10-3, 10-4, and 0.9, respectively. This phase lasts 20 epochs. The total computation time in this phase was around 1 hour.
2) Full-training Phase: We unfreeze the backbone model and update the entire network via SGD with momentum. The learning rate, weight decay, and momentum of SGD is set to 10-3, 10-4, and 0.9, respectively. At the epoch 40 of this phase, the learning rate is dropped by half. This phase lasts 80 epochs. After this phase, the model generated at the epoch with lowest validation loss is used as our final model. The total computation time in this phase was around 8 hours and 10 minutes.

IV. EXPERIMENTS AND RESULTS

In this section, we demonstrate the effectiveness of our proposed methods of polyp segmentation in colonoscopy images. We used 3 datasets to evaluate our methods and compared our results with those of state-of-the-art algorithms.

A. DATASET

In order to demonstrate the model performance of our proposed approach, we used three public available

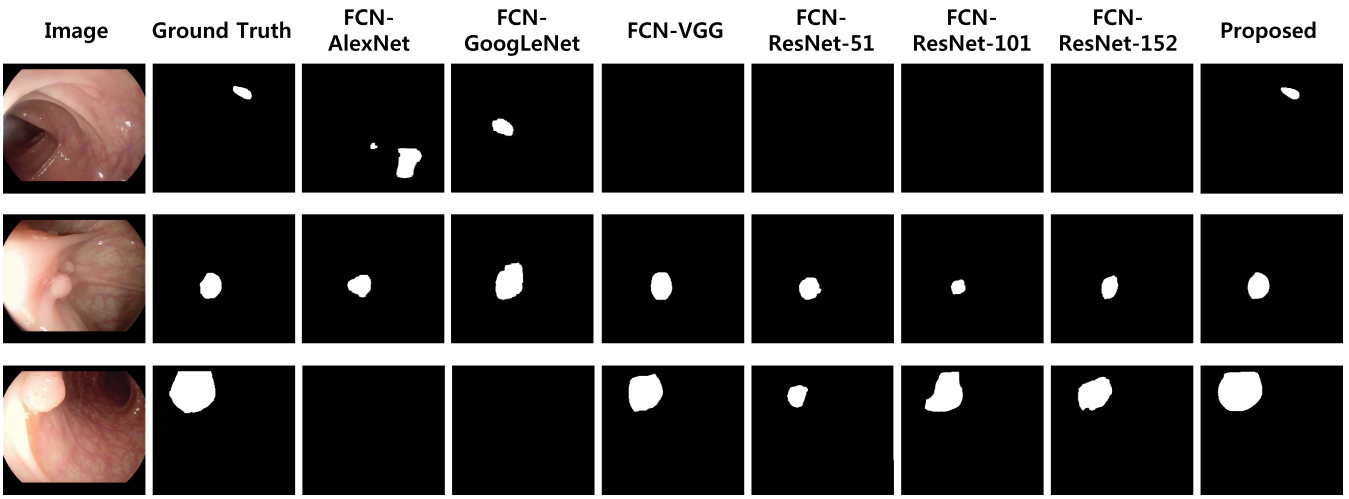

FIGURE 5. Example of three different segmentations produced by the six FCN networks and proposed method.

datasets: CVC-ClinicDB [27] and ETIS-Larib [10] from the MICCAI 2015 polyp detection challenge [9], and CVC-ColonDB [28]. There are similar image frames within the same colonoscopy dataset. If we use one dataset and divide that into training and testing sets, then exaggerated classification performance can be obtained. Therefore, for more reliable evaluation, we assign the above-mentioned different datasets into training and testing set separately as the recommendation of the MICCAI challenge guidelines: CVC-CLINIC for training and ETIS-Larib for testing. Furthermore, we also report results from other public available dataset (CVC-ColonDB) as a testing set. The datasets were obtained with different imaging systems and contain binary masks as the ground truths to indicate the location of the polyps for each image. All ground truths of polyp regions for these datasets were annotated by expert video endoscopists from the corresponding associated clinical institutions. We specifically used the following grouping of images for training (fine-tuning) and testing:

1) CVC-ClinicDB [27]: contains 612 polyp image frames with SD (standard definition) resolution of 388 × 284 pixels from 31 different colonoscopy video sequences with 31 unique polyps.
2) ETIS-Larib [10]: contains 196 polyp image frames with HD (high definition) resolution of 1225 × 966 pixels from 34 different colonoscopy video sequences. This dataset contains 44 different polyps with various sizes and appearances. At least one polyp existed in all 196 images, with the total number of polyps being 208.
3) CVC-ColonDB [28]: contains 379 polyp image frames with SD (standard definition) resolution of 574 × 500 pixels from 15 different colonoscopy video sequences with at least one polyp each.

B. PERFORMANCE METRICS

The proposed model was formulated to produce dense pixel-wise polyp segmentations. As such, we report results using three common segmentation evaluation metrics: 1) mean pixel precision (PR), 2) mean pixel recall (RC) and 3) interception over union (IU). If a pixel of polyp is correctly classified, it is counted as a true positive (TP). Every pixel segmented as polyp that falls outside of a polyp mask counts as a false positive (FP). Finally, every polyp pixel that has not been detected counts as a false negative (FN). The three evaluation metrics are calculated as follows.

$$PR = \frac{TP}{TP + FP} \tag{2}$$

$$RC = \frac{TP}{TP + FN} \tag{3}$$

$$IU = \frac{TP}{TP + FP + FN} \tag{4}$$

TABLE 1. Comparison with six FCN on the ETIS- LaribPolyp DB dataset.

Networks	PR (%)	RC (%)	IU (%)
FCN-AlexNet	27.87	35.54	15.7
FCN-GoogLeNet	25.83	29.82	12.29
FCN-VGG	70.23	54.2	44.06
FCN-ResNet-50	55.75	23.43	19.72
FCN-ResNet-101	63.26	53.88	41.35
FCN-ResNet-152	65.26	38.24	33.19
Proposed	**73.84**	**74.37**	**66.07**

C. RESULTS

Our results using ETIS- LaribPolypDB dataset are presented in the Table 1. Table 1 shows that our proposed model achieved the highest both precision and recall criterion among the other models on ETIS- LaribPolypDB dataset. When it comes to the model performance on CVC-ColonDB dataset, the dataset contains images all with polyps of different shapes which are annotated by clinicians. Table 2 shows that our approach totally outperformed FCN based methods on CVC-ColonDB dataset. This is because Mask R-CNNs

TABLE 2. Comparison with six FCN networks on the CVC-colon DB dataset.

Networks	PR (%)	RC (%)	IU (%)
FCN-AlexNet	40.3	20.71	15.77
FCN-GoogLeNet	37.46	12.93	12.71
FCN-VGG	76.06	60.46	54.01
FCN-ResNet-50	67.76	25.64	22.74
FCN-ResNet-101	73.85	50.73	46.23
FCN-ResNet-152	72.85	50.72	43.28
Proposed	**77.92**	**76.25**	**69.46**

used in our proposed approach can be viewed as a two-stage method that could reduce the negative influence of diverse size of polyp. In contrast, FCN based methods are one-stage segmentation methods that are sensitive to the size of the lesion. In addition, our ensemble model has the ability to aggregate the results from two different backbone structures, which can further improve the segmentation results.

Moreover, examples of three different segmentations produced by the six FCN networks and proposed method could be depicted in Figure 5. Figure 5 describes that our model can recognize the tumor boundary as much as possible what others could not do.

V. SUMMARY AND CONCLUDING REMARKS

We have proposed a transfer learning based ensemble model for colorectal polyp segmentation. The proposed framework consists of three elements: 1) data augmentation, 2) two Mask R-CNNs with different backbone structures (ResNet50 and ResNet101) pre-trained on the COCO dataset, and 3) the bitwise combination of two masks as our ensemble method. Our method is validated using well known dataset from MICCAI 2015 polyp detection challenge. Our experimental results demonstrated the superiority of our proposed method against state-of-the-art approaches for polyp segmentation.

Our research is still flawed, but we hope to try to break through existing research results in a variety of ways. In the colorectal polyp segmentation, a good classification algorithm and high quality data are complement each other. In the exploration of Mask R-CNN, we can explore other backbone structure. Besides, we will also continue to collect more high quality colorectal polyps dataset.

REFERENCES

[1] M. C. Parkin, F. J. Shin, and B. F. Forman, "Globocan 2008 v1.2, cancer incidence and mortality worldwide: Iarc cancerbase no. 10. International Agency for Research on Cancer, 2008.
[2] J. C. Van Rijn, J. B. Reitsma, J. Stoker, P. M. Bossuyt, S. J. Van Deventer, and E. Dekker, "Polyp miss rate determined by tandem colonoscopy: a systematic review," *Amer. J. Gastroenterol.*, vol. 101, no. 2, p. 343, Feb. 2006.
[3] D. Vázquez *et al.*, "A benchmark for endoluminal scene segmentation of colonoscopy images," *J. Healthcare Eng.*, vol. 2017, Jul. 2017, Art. no. 4037190.
[4] G. Tulum, B. Bolat, and O. Osman, "A CAD of fully automated colonic polyp detection for contrasted and non-contrasted CT scans," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 12, no. 4, pp. 627–644, Apr. 2017.
[5] Z. Wang, L. Li, J. Anderson, D. P. Harrington, and Z. Liang, "Computer-aided detection and diagnosis of colon polyps with morphological and texture features," *Proc. SPIE*, vol. 5370, pp. 972–980, May 2004.
[6] Y. Komeda *et al.*, "Computer-aided diagnosis based on convolutional neural network system for colorectal polyp classification: Preliminary experience," *Oncology*, vol. 93, pp. 30–34, Dec. 2017.
[7] S. Park, M. Lee, and N. Kwak, *Polyp Detection in Colonoscopy Videos Using Deeply-learned Hierarchical Features*. Seoul, South Korea: Seoul National Univ., 2015.
[8] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2980–2988.
[9] J. Bernal *et al.*, "Comparative validation of polyp detection methods in video colonoscopy: Results from the MICCAI 2015 endoscopic vision challenge," *IEEE Trans. Med. Imag.*, vol. 36, no. 6, pp. 1231–1249, Jun. 2017.
[10] J. Bernal, J. Sánchez, and F. Vilarino, "Towards automatic polyp detection with a polyp appearance model," *Pattern Recognit.*, vol. 45, no. 9, pp. 3166–3182, Sep. 2012.
[11] M. Ganz, X. Yang, and G. Slabaugh, "Automatic segmentation of polyps in colonoscopic narrow-band imaging data," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 8, pp. 2144–2151, Aug. 2012.
[12] P. Arbeláez, M. Maire, C. Fowlkes, and J. Malik, "Contour detection and hierarchical image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 5, pp. 898–916, May 2011.
[13] S. Hwang, J. Oh, W. Tavanapong, J. Wong, and P. C. de Groen, "Polyp detection in colonoscopy video using elliptical shape feature," in *Proc. IEEE Int. Conf. Image Process.*, Oct. 2007, pp. 465–468.
[14] N. Tajbakhsh, S. R. Gurudu, and J. Liang, "Automated polyp detection in colonoscopy videos using shape and context information," *IEEE Trans. Med. Imag.*, vol. 35, no. 2, pp. 630–644, Feb. 2016.
[15] N. Tajbakhsh, S. R. Gurudu, and J. Liang, "A classification-enhanced vote accumulation scheme for detecting colonic polyps," in *Abdominal Imaging. Computation and Clinical Applications* (Lecture Notes in Computer Science). Berlin, Germany: Springer-Verlag, 2013, pp. 53–62.
[16] E. Nasr-Esfahani *et al.*, "Melanoma detection by analysis of clinical images using convolutional neural network," in *Proc. 38th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Aug. 2016, pp. 1373–1376.
[17] N. Tajbakhsh *et al.*, "Convolutional neural networks for medical image analysis: Full training or fine tuning?" *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1299–1312, May 2016.
[18] L. C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam. (2018). "Encoder-decoder with atrous separable convolution for semantic image segmentation." [Online]. Available: https://arxiv.org/abs/1802.02611
[19] E. Ribeiro, A. Uhl, and M. Hafner, "Colonic polyp classification with convolutional neural networks," in *Proc. IEEE 29th Int. Symp. Comput.-Based Med. Syst. (CBMS)*, Jun. 2016, pp. 253–258.
[20] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 3431–3440.
[21] M. Akbari *et al.* (2018). "Polyp segmentation in colonoscopy images using fully convolutional network." [Online]. Available: https://arxiv.org/abs/1802.00368
[22] L. Zhang, S. Dolwani, and X. Ye, "Automated polyp segmentation in colonoscopy frames using fully convolutional neural network and textons," in *Communications in Computer and Information Science*. Cham, Switzerland: Springer, 2017, pp. 707–717.
[23] A. Krizhevsky, S. Ilya, and E. H. Geoffrey, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1106–1114.
[24] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proc. CVPR*, Jun. 2015, pp. 1–9.
[25] K. Simonyan and A. Zisserman. (2014). "Very deep convolutional networks for large-scale image recognition." [Online]. Available: https://arxiv.org/abs/1409.1556
[26] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.
[27] J. Bernal *et al.*, "WM-DOVA maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians," *Comput. Med. Imag. Graph.*, vol. 43, pp. 99–111, Jul. 2015.
[28] J. Silva, A. Histace, O. Romain, X. Dray, and B. Granado, "Toward embedded detection of polyps in WCE images for early diagnosis of colorectal cancer," *Int. J. Comput. Assist. Radiol. Surgery*, vol. 9, no. 2, pp. 283–293, 2014.

[29] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," In *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 91–99.
[30] G. Cheng, P. Zhou, and J. Han, "Learning rotation-invariant convolutional neural networks for object detection in VHR optical remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 12, pp. 7405–7415, Dec. 2016.
[31] W. Ouyang *et al.*, "DeepID-Net: Deformable deep convolutional neural networks for object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 2403–2412.
[32] Y. Zhang, K. Sohn, R. Villegas, G. Pan, and H. Lee, "Improving object detection with deep convolutional networks via Bayesian optimization and structured prediction," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 249–258.
[33] J. Zhang, P. Liu, F. Zhang, and Q. Song, "CloudNet: Ground-based cloud classification with deep convolutional neural network," *Geophys. Res. Lett.*, vol. 45, no. 16, pp. 8665–8672, Aug. 2018.
[34] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and L. Fei-Fei, "Large-scale video classification with convolutional neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 1725–1732.
[35] P. Sermanet, S. Chintala, and Y. LeCun, "Convolutional neural networks applied to house numbers digit classification," in *Proc. 21st Int. Conf. Pattern Recognit.*, Nov. 2012, pp. 3288–3291.
[36] J. M. Wolterink, T. Leiner, M. A. Viergever, and I. Išgum, "Automatic coronary calcium scoring in cardiac CT angiography using convolutional neural networks," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2015, pp. 589–596.
[37] H. R. Roth *et al.*, "DeepOrgan: Multi-level deep convolutional networks for automated pancreas segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent*, 2015, pp. 556–564.
[38] A. de Brebisson and G. Montana, "Deep neural networks for anatomical brain segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2015, pp. 20–28.
[39] P. Moeskops, M. A. Viergever, A. M. Mendrik, L. S. de Vries, M. J. N. L. Benders, and I. Išgum, "Automatic segmentation of MR brain images with a convolutional neural network," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1252–1261, May 2016.
[40] A. Prasoon, K. Petersen, C. Igel, F. Lauze, E. Dam, and M. Nielsen, "Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent*, Aug. 2013, pp. 246–253.
[41] J. Han *et al.*, "Representing and retrieving video shots in human-centric brain imaging space," *IEEE Trans. Image Process.*, vol. 22, no. 7, pp. 2723–2736, Jul. 2013.
[42] X. Ji *et al.*, "Retrieving video shots in semantic brain imaging space using manifold-ranking," in *Proc. 18th IEEE Int. Conf. Image Process.*, Sep. 2011, pp. 3633–3636.
[43] X. Hu *et al.*, "Bridging low-level features and high-level semantics via fMRI brain imaging for video classification," in *Proc. 18th ACM Int. Conf. Multimedia*, Oct. 2010, pp. 451–460.
[44] X. Hu, K. Li, J. Han, X. Hua, L. Guo, and T. Liu, "Bridging the semantic gap via functional brain imaging," *IEEE Trans. Multimedia*, vol. 14, no. 2, pp. 314–325, Apr. 2012.
[45] A. Karargyris and N. Bourbakis, "Identification of polyps in wireless capsule endoscopy videos using Log Gabor filters," in *Proc. IEEE/NIH Life Sci. Syst. Appl. Workshop*, Apr. 2009, pp. 143–147.
[46] Y. Jia, "Polyps auto-detection in wireless capsule endoscopy images using improved method based on image segmentation," in *Proc. IEEE Int. Conf. Robot. Biomimetics (ROBIO)*, Dec. 2015, pp. 1631–1636.
[47] H. Chen, X. Qi, L. Yu, and P. A. Heng, "DCAN: Deep contour-aware networks for accurate gland segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 2487–2496.

JAEYONG KANG received the Ph.D. degree in electrical engineering and computer science from the Gwangju Institute of Science and Technology, Gwangju, South Korea, in 2017. He is currently a Research Scientist with the Biomedical Research Institute, Seoul National University Hospital, Seoul, South Korea. His current research interests include deep learning, natural language processing, computer vision, agent-based information retrieval, semantic web, social media analysis, and recommender systems.

JEONGHWAN GWAK received the Ph.D. degree in machine learning and artificial intelligence from the Gwangju Institute of Science and Technology, Gwangju, South Korea, in 2014. From 2002 to 2007, he was with several companies and research institutes as a Researcher and the Chief Technician. From 2014 to 2016, he was a Postdoctoral Researcher with GIST and from 2016 to 2017, as a Research Professor. He is currently a Research Professor with the Biomedical Research Institute and with the Department of Radiology, Seoul National University Hospital, Seoul, South Korea, and the Director of the Applied Medical Machine Learning Laboratory. His current research interests include deep learning, computer vision, image and video processing, evolutionary computation, optimization, and relevant applications of medical and visual surveillance systems. He served as the Chairman and PC/TPC members in many artificial intelligence, machine learning, and computer vision conferences. He is an Associate Editor of the IEEE Access and the Guest Editor of IJCVR.