IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received August 30, 2019, accepted September 18, 2019, date of publication September 30, 2019, date of current version October 10, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2944676

Review on the Applications of Deep Learning in the Analysis of Gastrointestinal Endoscopy Images

WENJU DU[1,2], NINI RAO[1,2,3], DINGYUN LIU[1,2], HONGXIU JIANG[1,2], CHENGSI LUO[1,2], ZHENGWEN LI[1,2], TAO GAN[4], AND BING ZENG[5], (Fellow, IEEE)

[1]Center for Informational Biology, University of Electronic Science and Technology of China, Chengdu 610054, China
[2]Center for Information in Medicine, University of Electronic Science and Technology of China, Chengdu 610054, China
[3]Institute of Electronic and Information Engineering, University of Electronic Science and Technology of China (UESTC), Dongguan 523107, China
[4]Digestive Endoscopic Center, West China Hospital, Sichuan University, Chengdu 610017, China
[5]School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu 610054, China

Corresponding authors: Nini Rao (raonn@uestc.edu.cn) and Bing Zeng (eezeng@uestc.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61872405 and Grant 61720106004, in part by the Key Project of Natural Science Foundation of Guangdong Province under Grant 2016A030311040, in part by the Sichuan Science and Technology Support Program under Grant 2015SZ0191, in part by the Fundamental Research Funds for the Central Universities of China under Grant ZYGX2016J189, and in part by the Scientific Platform Improvement Project of UESTC.

ABSTRACT Gastrointestinal (GI) disease is one of the most common diseases and primarily examined by GI endoscopy. Recently, deep learning (DL), in particular convolutional neural networks (CNNs) have made achievements in GI endoscopy image analysis. This review focuses on the applications of DL methods in the analysis of GI images. We summarized and compared the latest published literature related to the common clinical GI diseases and covers the key applications of DL in GI image detection, classification, segmentation, recognition, location, and other tasks. At the end, we give a discussion on the challenges and the research directions of GI image analysis based on DL in the future.

INDEX TERMS Gastrointestinal disease, gastrointestinal endoscopy image, deep learning, analysis, comparison.

I. INTRODUCTION

GI disease is one of the most common diseases and commonly occurs in humans, resulting in one of the most important healthcare problems. According to the extent of the lesion, it can be roughly divided into benign GI diseases, precancerous lesion, early GI cancer and advanced GI cancer. Benign GI diseases such as ulcers, gastritis and bleedings will not deteriorate into cancers in short term. Precancerous GI lesions may deteriorate into early GI cancer or even advanced GI cancer, if not diagnosed and treated in time. The 2018 world cancer statistics [1] indicate that colorectal cancer, gastric (stomach) cancer (GC) and esophageal cancer are three main GI cancers. The highest incidence rates of colon cancer are found in western regions/countries such as Europe, Australia/New Zealand, and Northern America. Incidence rates of GC are markedly elevated in Eastern Asia, while the rates in the western countries are generally low. Esophageal cancer is common in several countries in Eastern Asia, Eastern and Southern Africa, with the highest rates in Eastern Asia. Clinical data suggest that the 5-year survival rate of GC remains low (between 23% and 27%) [2], while the 5-year survival rate of advanced gastric cancer (AGC), especially TNM stage IV cancer, is only 4% [3]. However, the 5-year survival rate of early gastric cancer (EGC) can be as high as 95% [4]. Therefore, the earlier the detection and active intervention of GC, the higher the survival rate of patients, with even the potential of fully recovery. The accurate detection and diagnosis of precancerous lesions and early cancer of GI are crucial to prevent GI diseases from developing into advanced cancer.

Currently, the examination and diagnosis of GI diseases mainly rely on endoscopy [4], [5]. This technique is a meth-od to noninvasively deliver a pathological diagnosis of living tissue. GI endoscopy includes gastroscopy, colonoscopy and wireless capsule endoscopy (WCE), where images captured by these three endoscopies are shown in Fig. 1. Usually, gastroscopy is used to examine abnormalities of the upper

The associate editor coordinating the review of this manuscript and approving it for publication was Jihwan P. Choi.

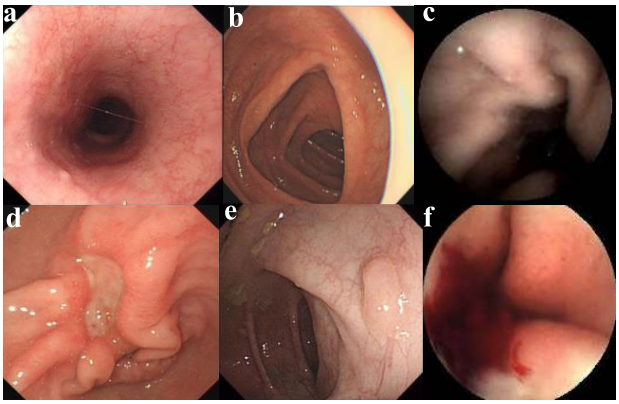


FIGURE 1. Examples of GI endoscopic images. (a) Gastroscopic image of normal esophagus, (b) colonscopic image of normal colorectal, (c) WCE image of normal small intestinal, (d) gastroscopic image of stomach with ulcer, (e) colonscopic image of colorectal with polyp, (f) WCE image of small intestinal with hemorrhage.

GI, colonoscopy is used to check the lower GI and WCE is main-ly used to detect lesions in the small intestine between the upper and lower GI. WCE, a new type of micro-digestive endoscopy, involves a small device that is swallowed by patients, passes through the overall GI and is then discharged from the anus, examining the entire GI along the way within the battery life time of about 8 hours. The number of GI images produced by each endoscopic examination is very large, especially in WCE, which can produce 50,000 to 120,000 images during one examination. The reading of a large amount of endoscopic image data has exceeded the limitation of human concentration, thus easily resulting in misdiagnosis and a decrease of diagnostic accuracy. In addition, the diagnostic results may be controversial, due to different experiences of doctors. In fact, only a small number of GI images would contain GI lesions. Selecting a small number of crucial lesion images from a large number of endoscopic images is a time-consuming and laboriously inefficient task for doctors. To improve the efficiency and diagnosis accuracy, some computer aided diagnosis (CAD) systems have been developed, which automatically select, identify and classify lesion images, and provide an objective reference to doctors. These CAD systems could not only reduce the burden of doctors, but also improve the diagnostic efficiency and brings a great help to doctor.

The traditional framework of a CAD system consists of a feature extraction and a classifier based on machine learning (ML) methods. First, an artificially designed algorithm is used to extract image features such as color and texture; then, these extracted features are sent to a classifier such as a Support Vector Machine (SVM) [6]–[9]. For instance, Liu *et al*. [7] designed a joint diagonalisation principal component analysis algorithm to extract features of endoscopic images; then, these features were sent into a SVM to be classified into two categories: abnormal and normal images. A comparison between hand-craft feature based SVM and CNNs-based DL for colon polyp detection was performed by Shin and Balasingham [10]; the results indicated that the CNNs-based DL method performed better. The Endoscopic Vision Challenge results of 2015 Medical Image Computing and Computer Assisted Intervention (MICCAI) demonstrated that a method based on DL is the state-of-the-art [11]. Besides, in two papers which studied the detection of intestinal hookworms, the accuracy of the DL method [12] was found to be approximately 10.3% higher than that of the artificial feature extraction method from [13] using the same database.

Recently, DL has achieved a great success in the field of computer vision. In certain cases, its object recognition accuracy can even surpass that of human beings. In particular, CNNs have achieved very good results in different image processing tasks [14]. CNNs first appeared in1980 [15], and Lo *et al.* [16] first applied CNNs to lung nodule detection in 1995. The first successful application of CNNs was LeNet which was used for digital handwriting recognition in 1998 [17]. Although these studies highlighted the initially great successes of the application of CNNs, the usefulness of this kind of network seemed to be halted because of the limited computational power at that time. Alternatively, scholars preferred to choose other methods, such as artificial feature extraction methods and so on. CNN was gradually forgotten over the next decade. It is not until 2012 that AlexNet [14] was proposed and won the ImageNet Large-Scale Visual Recognition Challenge (ILSVR-C), with the top-5 error rate around 10% higher than the second place. Since then, CNNs have become increasingly popular. Subsequently, DL technique was quickly applied in various fields, and an increasing number of scholars have begun to explore the applications of DL methods in medical image analysis [18]–[20] and have obtained quite well results. For instance, in [18], the classification accuracy of skin cancer was found to be close to that obtained by dermatologist. In recent years, DL technique has gradually been applied to the image processing of GI, and several papers have been published as pioneering works in this field [21]–[25]. In the latest published literature [21], Shin *et al*. presented the first successful case of applying the DL technique GI polyp detection, while the authors of [22] realized a real-time detection of colorectal polyps. Meanwhile, a 3D-FCN (fully convolutional network) was first used to identify polyps in a colonoscopic video by Yu *et al.* [25], and Jia and Meng [23] were the first to explore the automatic detection of intestinal bleeding, and the authors of [24] tried to classify EGC using some DL methods.

To the best of our knowledge, this is the first review on the applications of DL methods in the analysis of GI images, and we believe that it can provide an important reference for researchers in this field. Other reviews about the applic-ations of DL methods in medical image analysis such as [26]–[29] only involve few works related to GI image analysis. The rest of the paper is organized as follows: Part II will provide an overview of DL methods. Part III will intro-duce the application of DL in GI endoscopic image analysis. Part IV provides a comprehensive overview of the literature cited in this review

and discuss several issues encountered in the application of DL methods in GI image analysis. Part V, briefly summarizes some significant research directions for the future works.

II. OVERVIEW OF DEEP LEARNING METHODS

This section provides an introduction of DL methods [30], which are a branch of ML. Both of DL and ML belong to artificial intelligence (AI). DL architectures refer to neural networks with large amounts of hidden layers. Recently, DL methods have been regarded as the most advanced AI techniques by virtue of their state-of-the-art performances, especially deep convolutional neural networks (DCNNs) have brought breakthroughs in image processing.

The training of DL methods is usually divided into two categories: supervised learning and unsupervised learning. The commonly used DL architectures in GI image analysis are trained in a supervised manner with labeled data. As presented in Table 1 that almost all the literature related to deep networks used in GI image analysis are based on CNN (supervised learning), while only 2 papers apply other networks such as artificial neural network (ANN) and deep neural network (DNN). Next, we will give a detailed introduction of CNN, and a brief introduction of other DL architectures used in GI image analysis.

TABLE 1. Sumarry of deep architectures used in gastrointestinal image analysis.

		Deep Architectures	N.
Supervised	NN	ANN, DNN	2
	CNN	LeNet, AelxNet, GoogLeNet, VGGNet, ResNet, InceptionResNet, SSD, FCN, Fast R-CNN, Faster R-CNN, DeepLab, SegNet	44
Unsupervised		GAN	1

N.=the number of papers

A. CONVOLUTIONAL NEURAL NETWORK

The working principle of CNN can be illustrated by two steps. Firstly, the network is trained over a given labeled dataset and the multiscale features are extracted. Secondly, based on the features extracted by the first step, classification is performed. CNN consists of several important components, including convolutional layers, activation functions, pooling layers and fully connected layers. A simple CNN usually consists of several of these layers, while some very deep CNN models could include hundreds of layers. For instance, one version of the current popular ResNet consists of 152 layers.

The convolutional layer is a crucial component of CNN, and the neurons in the convolutional layer are sensitive to every small piece of the input images. In the terminology of CNN, the first parameter of the convolution is usually called input, the second parameter is called the kernel function, and the output sometimes is called feature map, as shown in equation (1), where x is the input, ω is kernel function, s(t) denotes the output feature map. The definition of two dimension (2D) convolution operation is shown as equation (2), where I is the input, K denotes a 2D kernel function.

$$s(t) = (x * \omega)(t) = \sum_{a=-\infty}^{\infty} x(a)\omega(t-a) \tag{1}$$

$$S(i,j) = (I * K)(i,j) = \sum_{m}\sum_{n} I(m,n)K(i-m, j-n) \tag{2}$$

The selection of the activation function for a CNN is very important. Currently, a rectified linear unit (ReLU) is the preferred activation function, which is defined as follows:

$$f(x) = \max(0, x) \tag{3}$$

The pooling layer could reduce computational costs by computing the overall statistical characteristic of the adjacent rectangular region of a location to replace the output of the convolutional layer at that region. For example, a max-pooling layer, the most commonly used kind of pooling layer, computes the maximum value of the adjacent rectangular area. Except for max-pooling, there are many other pooling layers, such as average-pooling and L2-norm pooling.

The last layers of CNN are the fully connected layers, in which each neuron in the layer is connected to each neuron in the next layer. The output of the previous layers could be sent to fully connected layer as an input, and a probability score for each class to which the input image could be assigned is computed. The class with the highest score is the final classification result of the input image. In short, the fully connected layer combines the most prominent features of the image to infer the category of an image.

An example of classifying GI image by CNN is shown in Fig. 2. First, the features of the input image are extracted by convolutional layers, activate functions and pooling layers. Then, the output feature map is sent to the fully connected layers, and the prediction probability scores (between 0-1) for Lesion 1, Lesion 2, Lesion 3 and Normal category are computed out. In this example, the prediction probability scores of Lesion 1, Lesion 2 and Lesion 3 are very small, while the probability score of Normal class is 0.96. Thus, the input image is classified as normal category.

B. SUPERVISED DEEP LEARNING ARCHITIECTURES

In this section we give an overview of the commonly used DL architectures based on the supervised manner in GI image analysis.

1) CLASSIFICAITON ARCHITECYURES

The most popular deep models used in GI image classification are LeNet, AlexNet, VGGNet, GoogleNet, ResNet and so on.

LeNet [17] and AlexNet [14] are relatively shallow, they explore kernels with large receptive fields in layers close to the input and smaller kernels close to the output. One difference between these two architectures is that AlexNet us es ReLU unit instead of the hyperbolic tangent as the activation function, which is the mostly used nowadays. VGGNet (also

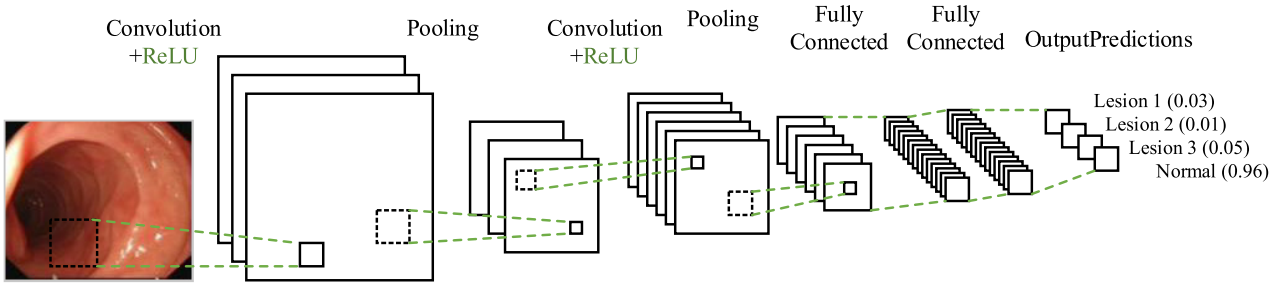


FIGURE 2. A simple example of GI image classification using CNNs. The features were extracted by convolution layers, and then sent to fully connected layers. The predicted classification results were given out by fully connected layers.

called OxfordNet) [31] was proposed by Simonyan *et al.*, which is also relatively shallow and consists 16-19 layers.

Nowadays, there is a preference for deeper models and smaller kernels instead of a single layer and kernels with large receptive field, because a smaller function means fewer parameters, such as GoogleNet and ResNet. Szegedy *et al*. [32] proposed GoogLeNet (also called Inception), which introduced inception block that has been shown to be able to achieve very good performance at low computational cost [33]. ResNet [34] consists of the ResNet-blocks which only learns the residual function with reference to the layer inputs, rather than learning function without reference. The experiment evidences showed that these residual networks are easier to be optimized and can gain accuracy from increased depth. In other words, even deeper architectures can also be trained effectively.

Since 2012, the performance of ILSVRC became a benchmark. Squeeze-and-Excitation Networks [35] won the last ILSVRC of 2017, which has not yet been used in GI image analysis. The performances of these popular classification architectures on ImageNet database are shown in Fig. 3, where, the Top-5 error rate is that the fraction of test images for which the correct label is not among the five labels considered most probable by the model. We can see from Fig. 3 that the Top-5 error rate of deep models on ILSVRC keeps to be smaller year by year, but the accuracy seems to get saturated. It is not sure that the small increases in performance could be attributed to more sophisticated architectures of a deep network. Additionally, GI image is different from nature images. Therefore, the respective shallow and simple networks such as AlexNet, VGG are still popular for GI image analysis.

2) DETECTION ARCHITECYURES

Currently, detection by DL methods is a common task in GI image analysis. There are three object dection methods based on CNNs: single shot multibox detection (SSD) [36], fast region-based convolutional neural network (Fast R-CNN) [37], and Faster R-CNN [38], which are popularly used in the GI image analysis. The SSD method transforms object detection into an end-to-end target detection for regression problems. Fast R-CNN and Faster R-CNN combined region proposal algorithm and CNN classification together.

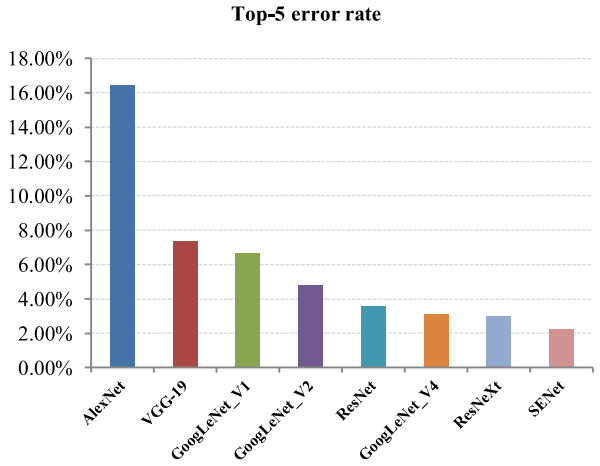


FIGURE 3. The top-5 error rates of classification of the current popular deep networks on ImageNet. It can be seen that the performances of these latest deep architectures has been improved little.

3) SEGMENTATION ARCHITECTURES

The segmentation of a GI image generally refer to semantic segmentation. FCN [39], DeepLab [40] and SegNet [41] are semantic image segmentation (also called pixel-wise classification) architectures, and are trained in an end-to-end manner.

Since all layers in FCN are convolutional layers, it is named as fully convolutional networks. Compared with the traditional segmentation method based on CNNs, there are two distinct advantages in FCN: (1) it is more flexible since the input images of FCN can be of any size, (2) it is more effective since it uses pixel blocks and avoids the problems of repeated storage and convolution calculation.

The major contributions of DeepLab are as follows: (1) Speed: it accepts atrous convolution algorithm. (2) Accuracy: they obtain the state-of-the-art result. (3) Simplicity: their system is composed of DCNNs and conditional random fields (CRFs). (4) Atrous spatial pyramid pooling (ASPP) is introduced in DeepLab_V2 and the later versions.

SegNet shares the same property with U-Net [42], which has a pair of encoder and corresponding decoder networks. The highlight of SegNet is that the max-pooling indices are transferred to the decoder, which improves the segmentation resolution. Both of them are effective semantic image segmentation architectures.

C. UNSUPERVISED DEEP LEARNING ARCHITIECTURE

Generative adversarial network (GAN) [43] is an unsupervised architectures, which holds promise for the GI image analysis task. GAN is composed of two simultaneously trained and competing models: a generative model G that captures the data distribution, and a discriminative model D that estimates the probability that a sample come from the training data rather than G. During the training procedure, G tries to maximize the probability of D making mistake. This model is also described as a minimax two-player game. At the end, there is a unique solution, where G recovers the training data distribution and D equals to 1/2 everywhere. Both G and D can be trained with back propagation, and without unrolled approximate inference and Markov chains.

D. OTHER NETWORKS

In addition to the DL networks used in GI image analysis, there are many other efficient networks such as recurrent neural networks (RNNs), graph neural networks (GNNs) [44], principle component analysis network (PCANet) [45] and canonical correlation analysis network (CCANet) [46] that have not yet been used in GI image analysis at present.

RNNs were developed for discrete sequence analysis. They have been used in other medical images analysis tasks such as Tissue segmentation [47]. GNNs were first proposed in 2009, which apply the existing neural network methods for processing data represented in a graph domain. GNNs have been widely applied to natural or other images processing tasks, but there are no related papers applying this method to GI images and other medical images.

RNNs could map input sequences to output sequences [48], and are more capable in serialized data processing. For example, the work in [49] combines RNNs and CNNs together, which allows the processing of all contextual information regardless of image size. As GNNs endows the DL model with some causal reasoning ability, makes them could deal with rich relation information among elements which could be useful in diseases classification [50].

PCANet and CCANet are effective networks and have been used in nature image classification. One difference between them is that PCANet can only handle data represented as one-view features and CCANet could classify images represented by two-view features.

In a world, RNNs, GNNs, PCANet and CCANet are all promising in the GI image analysis task in the future.

E. TRANSFER LEARNING METHODS

Training a deep network from scratch needs a large number of labeled data, and the training and optimizing process of the network is usually very time consuming. Collecting a large number of GI image and annotating the corresponding labels by experts are also tough and error prone tasks. Hence, most of GI image analysis tasks based on DL methods adopt transfer learning approach, which can reduce the need of a deep network for training data. In the transfer learning terminology, the deep model trained on large image dataset (such as ImageNet) is called pre-trained model.

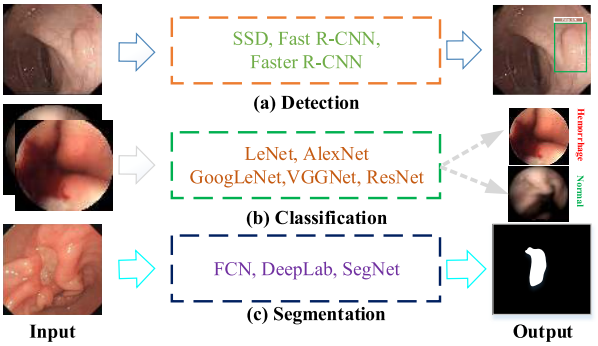


FIGURE 4. The illustrations of three main GI image analysis tasks: (a) Detection, (b) classification, (c) segmentation.

One transfer learning method is the feature extractor. The CNN layers of pre-trained model are used as feature extractor, and the fully connected layers of the pre-trained model are replaced by traditional classifier, such as linear classifier SVM. The GI image analysis tasks with a small number of samples usually choose this transfer learning method.

Another transfer learning method is the so-called fine-tuning. The input layer of pre-trained model is replaced and trained by new data. One can choose to fine-tune several layers or all layers of the pre-trained deep model. Typically, the previous layers of a deep network extract the generic features of the images (such as edge, color), which are useful for many tasks. The latter layers extract features related to a particular task, so fine-tuning method often only fine-tune the latter layers.

In addition, the other transfer learning method is parameter sharing. The parameters of a pre-trained deep network are loaded as the initialization parameters and trained with new data again, which can speed up the training process. The new trained model shares the same network and parameters with the pre-trained model. This transfer learning method usually requires a large training dataset.

III. APPLICATION OF DEEP LEARNING IN THE ANALYSIS OF COMMON GASTROINTESTINAL DISEASES

The applications of the DL methods in GI image analysis tasks include image detection, classification, segmentation, recognition, location, and a few other application tasks. The first three tasks are illustrated in Fig. 4. At present, the involved GI diseases mainly included polyps, hemorrhages, cancers, with some forays into the detection of gastritis and hookworms.

A. POLYPS

Polyps, one of the most common symptoms of GI, can be divided into hyperplastic polyps and adenomatous polyps according to their probability of progression into cancer. The former can be considered as benign polyps [51], the cancerous rate of which is relatively low, whereas the latter

has a higher cancerous rate, according to the clinical experiences [52]. Accurate identification and classification of hyper-plastic and adenomatous polyps can provide an objective reference to doctors, significantly improve the doctor's diagnosis efficiency, and effectively prevent the occurrence of early cancer.

1) DETECTION AND CLASSIFICATION

The detection and classification of colorectal polyps by DL methods have been explored in several works. The authors of [21] were the first to use a Faster R-CNN combined with a CNN model (Inception ResNet) to detect colonic polyps in images and videos. The novelty of this research is the proposed post learning that could effectively reduce the number of false positives (FPs) samples. After trying several data augmentation methods, their detection precision reaches 91.4%, but the mean detection time is about 0.39 second per frame and need to be further improved. Similarly, the classification of colonic polyps by several different CNNs was explored by the authors of [53]. Each of the polyp images was divided into a number of sub-images, which could increase the number of training datasets and reduce the computational complexity of the network. To improve the stability of DCNN model identifying polyps in complex environments, Karnes *et al*. [54] used a database of both white light and narrow-band imaging (NBI) colonoscopic images to train a CNN for classifying the image samples into polyps and normal tissues. It is very difficult to evaluate the performance of the model in different environments. Bernal *et al*. [11] conducted a unified evaluation experiment on the eight polyp detection methods of the MICCAI 2015 Endoscopic Vision Challenge (one method based on artificial feature extraction, four based on CNNs, and three hybrid methods. The results showed that the DL methods present the state-of-the-art, and hybrid methods can improve overall performance.

The research works aforementioned can only detect whether the images contain polyps or not. If the detected polyps could further be classified according to the rate at which they could develop into tumors, the procedure would be even more beneficial to both doctors and patients. The automatic detection and detailed classification of colorectal polyps based on the DL methods was explored by Zhang *et al*. [52]. They studied the transfer learning of different DCNNs and automatically classified colonoscopic images into hyperplasic polyps, adenomatous polyps, and normal images. The precision of their method was 87.3%, which is similar to the 86.4% precision from a physician. In the mean time, the recall rate and accuracy from the DL method were 87.6% and 85.9%, respectively, which were much higher than 77.0% and 74.3% achieved by physician. A similar study can be found in [55], which explored 6 different pre-trained deep networks by transfer leaning and training from scratch methods to classify colorectal polyps into hyperplastic polyps, adenomatous polyps and malignant polyps. A CAD system based on the transfer learning of the DNN was designed by Chen *et al*. [56], which classified diminutive colorectal polyps into three sub-types: hyperplastic polys, adenomatous polyps, and normal tissues.

Byrne [22] designed a system for the real-time assessment of polyp subtypes in colonoscopic videos based on DL methods. The data used in this research only included the colonoscopic videos of NBI with a period of only 50 ms between the two frames, which is a difficult task as a high request for the speed of the image recognition.

Except for the real-time video polyp detection system based on 2D-CNNs, some scholars have also explored a real-time video polyp detection system based on 3D-CNNs. The 3D-CNNs can better encode the video spatial information and learn more spatial features. Yu *et al.* [25] are pioneers in exploring a novel online and offline DL frameworks based on 3D-FCNs to automatically detect polyps in colonoscopic videos which can reduce the number of FPs. In a video subtest where each frame contains polyps, the method reached 0 FPs and 100% precision. However, the test video data where each frame contains polyps is unlikely to occur in actual clinical practice. Therefore, there is still some room for improvement of this method. Similarly, Tajbakhsh *et al*. [57] used a 3-way image representation and CNNs to detect polyps automatically in a colonoscopic video. In this study, the three characteristics: color and texture clues, shape in context, and temporal features of the polyp image, were extracted and then sent into the three CNNs for training, respectively. This method could provide more precise locations of the polyps, with only 0.002 FPs per frame at a sensitivity of 50%. There are differences between the two studies above. One uses a 3D DL frameworks and can learn more spatial features with encoded 3D information [25], while the other applies a 3-way images [57] to simultaneously train three CNNs, and learn three characteristics from lesion, respectively.

2) SEGMENTATION

Most of the researchers focus on polyp detection and classification, few of them have tried segmentation. Xiao *et al*. [58] attempted to use a DNN called DeepLab_v3 to detect polyps in colonscopic images. As the large structure of DeepLab_v3, the location of polyps may not be saved and transmitted effectively. To avoid this problem, the authors combined long short-term memory (LSTM) network with DeepLab_v3 in parallel to augment the location signal of polyps. They found a quite satisfactory results: mean intersection over union (mIOU) of 93.21% and the average computing time of 0.023 second per image. In summary, CNNs have been applied in the detection, classification, segmentation of colorectal polyps and have achieved quite well results. An overview of papers related to application of DL techniques on polyps is listed in Table 2. As depicted in Table 2, most papers focus on the polyps classification and detection tasks. In other words, more papers on other tasks should be encouraged. Fig. 5 shows the comparisons of detection and classification accuracies of several papers related to polyp detection. We can get an overview of the performances of these approaches in different references from Fig. 5 directly.

TABLE 2. Overview of papers using DL techniques for polyp image analysis.

Reference	Networks	Remarks
	Detection	
Shin *et al.*[21]	Faster R-CNN, Inception ResNet	Used region based DCNN and post learning method, detect polyps in video and images
Karnes *et al.*[54]	CNN	Detected polyps in colonoscopic images with both white light and NBI images
Bernal *et al.*[11]	CNN and other methods	Comparative evaluation of polyp detection methods of MICCAI 2015
Yu *et al.*[25]	3D-FCN	Novel online and offline methods based on 3D -FCN for polyp detection in video
Tajbakhsh *et al.*[57]	CNN	Using a 3-way image representation, could provide more accuracy location information
	Classification	
Ribeiro *et al.*[53]	CNN	Divided a image into sub-images to increase the size of database
Zhang *et al.* [52]	CNN	Containing both polyp detection and classification
Ribeiro *et al.*[55]	CNN	Investigated 6 different deep models
Chen *et al.*[56]	DNN	Classified diminutive polyps of high-quality, NBI colonoscopic images
Byrne *et al.*[22]	CNN	Real-time polyps classification in endoscopic video images
	Segmentation	
Xiao *et al.*[58]	DeepLab_V3	Combined LSTM networks and DeepLab in parallel

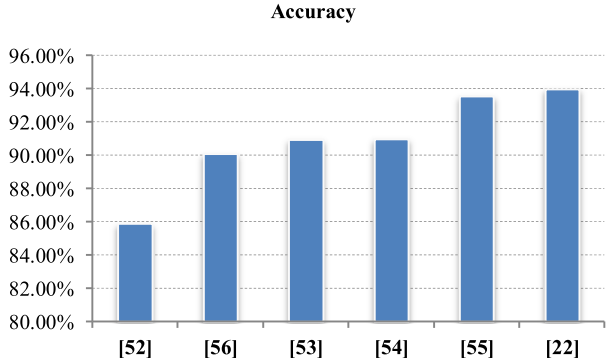


FIGURE 5. Comparison of the accuracies in part of polyp detection and classification references.

The results suggest that the DL architectures perform well on their data, as data scale and model vary from one paper to another.

It is worth mentioning that some scholars have tried to perform real-time detection, identification and classification of polyps in videos based on 2D-CNNs and 3D-CNNs. Compared to the image-based detection, video-based real- time detection is more helpful for doctors' aided diagnoses and endoscopic surgeries.

B. HEMORRHAGES

1) DETECTION AND CLASSIFICATION

Intestinal chronic hemorrhage is associated with GI diseases caused by unknown reasons [61]. Detection of GI hemorrhage (Fig. 4 (b)) is important for preventing their further deterioration and potential conversion into cancers.

A CNN containing 8 layers was designed by Jia and Meng [23] and trained on a dataset containing 10,000 WCE images, to detect GI bleeding (also called hemorrhage). Compared with traditional methods based on manual feature extraction, their methods performed better on all evaluation indicators. Another method for the detection of intestinal hemorrhage was explored by Li *et al.* [59]. This method is based on the transfer learning of several DL models, which involve the traditional LeNet and several state-of-the-art networks such as AlexNet, GoogleNet, and VGGNet. The authors also explored the effect of data augmentation on detecting accuracy. Jia and Meng [60] proposed a method that integrates manually extracted features and CNN layers extracted features, and sent them into the fully connected layer of CNN for classification. The results showed that although the training dataset was limited, the method achieved a precision of 94.79%, which was higher than that of other methods.

TABLE 3. Summary of the performacne of hemorrhage detection.

Ref.	Recall	Precision	F1-score	Data(hemorrhage+normal)
[23]	0.9920	0.9990	0.9955	2875+7150
[59]	0.9917	1.0000	0.9887	1300+40000
[60]	0.9100	0.9479	0.9285	300+1200

Ref. = reference

These research works all focused on WCE bleeding detection (classification) and the performances were summarized in Table 3. The performances of [23] and [59] are almost the same. It is difficult to distinguish which one is better. But the positive sample in [23] contains both active and inactive bleeding regions, maybe it is more challenging. In contrast, the results of [60] are slightly inferior because of the small scale of dataset. Considering the unbalance dataset (Table 3 last column) problem, these results are all quite satisfactory.

2) SEGMENTATION

Segmentation of hemorrhage lesions in WCE images was recently studied by three researchers. Jia and Meng [62] presented a method for automatic segmentation of hemorrhage region in WCE images. First, an SVM classifier was used to roughly divide the images into active bleeding group and inactive bleeding group according to the color features. Then, FCNs was applied to mark the two kinds of hemorrhage regions and achieved segmentation.

GI angiectasia is with inherent risk for bleeding. Leenhardt *et al.* [63] tried to perform a CNN-based semantic segmentation for deep feature extraction and classification of

TABLE 4. Comparisons of papers using dl method for hemorrhage segmentation.

Ref.	Network	Remarks	performances
[62]	FCNs	First classify into active and inactive lesions, then perform segmentation separately	mIU:0.7750, Fw IU:0.7854, mACC:0.8691, pACC:0.8796 (Active) mIU:0.7724, Fw IU:0.9848, mACC:0.8030, pACC:0.9917 (In-active)
[63]	CNN	A hand-crafted colorimetric segmentation approach combined with a CNN-based method for feature extraction	SE:100%, SP:96%, of 96%, PPV:100%, NPV:100%
[64]	GANs	Performing hand crafted and DL based methods for angiectasia segmentation both pixel-wise and frame-wise	SE:88%, SP:99.9% (pixel-wise location) SE:98%, SP:100% (frame-wise detection)
[65]	SegNet	Experiment on different color planes	gACC:94.42%, mACC:87.48%, MIoU:75.63%, wIoU:90.96%
[66]	CNN ANN	Focus on the problem of simplification the segmentation networks and reduce the computational complex	AUC:0.985

IU (IoU) stands for the region intersection over union, mIU = mean IU, mIoU = mIoU, wIoU = weight IoU, ACC = accuracy, mACC = mean accuracy, gACC = global accuracy, Fw = frequency weight, pACC = pixel accuracy, H = hemorrhage, SE = sensitivity, SP = speccifitic, PPV = positive predictive value, NPV = negative prediction value,

small intestine static frames to detect angiectasia. The authors of [64] explored a segmentation method based on GANs, which is able to mark the angiectasia in a given WCE video frame with pixel-wise accuracy. Ghosh *et al*. [65] developed a semantic segmentation approach based on SegNet for bleeding region detection in WCE images. The authors further tested the approach on different color planes and the best performance is achieved by using the hue saturation value (HSV) of color space. In [66], the authors investigated the problem of simplification of neural networks for automatic bleeding region segmentation in WCE images. The results showed that the simplification method on neural network and CNN structure could significantly reduce the burden of computational operation, which will reduce the detection time, especially for large number of WCE images. It has a great advantage for images retrieval in large dataset and endoscopic video abstract.

In general, the main problem of DL method in the hemorrhage analysis is the unbalanced data of abnormal and normal samples (as shown in Table 3, last column), which is also the case in other GI image analysis tasks. The problem of unbalanced sample is easy to cause poor generalization ability and over-fitting of the model. This is a stumbling block for the application of DL methods in the GI analysis. The comparisons of hemorrhage segmentation tasks between different references are listed in Table 4. Besides, the classification of hemorrhage's subtypes has still not been investigated in the existing researches. In short, further studies on GI hemorrhage detection, classification and segmentation based on DL methods are still needed.

C. GASTROINTESTINAL CANCER

The 5-year survival rate of EGC is up to 95%, which is much higher than that of AGC, especially the TNM stage'IV. However, early cancer may deteriorate into advanced cancer if it could not be timely treated. Hence, the detection and localization of early cancer can help doctors to improve the diagnosis accuracy and reduce misdiagnosis rate, which significantly improves the survival and cure rate of patients.

For some EGC and gastric ulcers, doctors with many years of experience still may not be able to distinguish these lesions [67]. As it is difficult to further improve the accuracy of conventional detection methods [10]–[13], [24], the application of DL methods in GI early cancer detection has been recently explored by some scholars. The authors of [24] pioneered the application of three efficient DCNN models, VGG16, InceptionV3 and InceptionResNetV2. They classified magnification endoscopy with narrow-band imaging (M-NBI) images into EGC and normal gastric images. Among their experimental results, the InceptionV3 network with fine-tuning transfer learning manner produced the best results with the values of evaluation parameters: accuracy, sensitivity and specificity were 0.985, 0.981 and 0.989, respectively. In addition, the authors also explored the effects of four different factors (training dataset, basic CNN architectures, fine-turned layers number and input image size) on transfer learning and compared the results of their method with those using traditional manual features, which provides a valuable reference for us.

Hirasawa *et al*. [68] performed more researches on the application of DL approach to detect EGC. They utilized a CNN framework called SSD to detect and locate EGC lesions in endoscopic images. The lesions in the output image were marked by rectangular windows with an annotation of the disease name and the probability that the lesion belongs to this disease (Fig. 4 (a)). Although the overall sensitivity of the method reached 92.2%, the missed lesions were all superficially depressed or belonged to intra-mucosal cancers (these kinds of lesions are more likely to be misdiagnosed even by experienced clinicians), which suggests that the method did not solve the clinical problem of diagnosing of these lesions completely. Additionally, nearly half of the FPs were gatritis lesions with irregular mucosal surfaces or color tone changes. Hence, the performance of this method is still with some room for improvement. The authors of [69] also designed a system based on SSD for the diagnosis of superficial (early cancer) and advanced esophageal cancer. The diagnostic accuracy and sensitivity of their method were both 98%, and this method even detected 2 more lesion regions missed by a previous examination. Riel *et al*. [70] designed a transfer learning method to automatically detect early esophageal cancer. They applied four pre-trained CNN models as feature extractors and then used traditional classifier, such as linear

SVM or softmax, to replace the fully connected layers. At the end of this study, the authors designed a method based on a sliding window to obtain a coarse-grained annotation of any possible cancerous lesions. The area under receiver operating characteristic (ROC) curve (AUC) of this approach was 0.92. Also, it allows for both near real-time prediction and annotation at 2 fps (4 frames /second).

For many kinds of cancers, a pathological diagnosis remains the gold standard. At present, the pathological diagnosis of tissue biopsies mainly relies on the experience of clinicians and is susceptible to subjective facts. Sometimes, it is difficult for human eyes to distinguish the subtle differences between benign and malignant tumors. However, DL networks are competent in solving this problem. Therefore, some scholars have started utilizing the DL methods to analyze these tissue biopsy images of EGCs. A new ResNet containing 50 layers was proposed by Liu *et al*. [71] to identify gastric pathology images (slices), and the F-score of this method was 96%. Similar to Liu's work, the authors of [72] proposed a network called GastricNet to detect gastric slices, and the classification accuracy of this method reached 100%. For the same goal, Qu *et al*. [73] utilized low cost medium-level datasets and a transfer learning method based on a stepwise fine-tuning scheme was used to train a deep network, which allows the network to understand a pathologic image from a pathologist's perspective.

The invasion depth of EGC is vital important as it determines whether an endoscopic resection could be performed or not for patients. The authors of [74] constructed a CNN based CAD system to determine the invasion depth of GC based on GI image and screened patients for endoscopic resection. The CNN based CAD system was trained in a transfer learning manner and the ResNet50 was chosen as a pre-trained architectures. The CNN based CAD system could distinguished EGC from deeper sub-mucosal invasion and minimized overestimation of invasion depth, which could reduce unnecessary gastrectomy and relieve the pain of patients. This system could provide an objective reference to doctors when they make decision on the treatment strategy of the GC patients.

Precancerous lesions may deteriorate into early GI cancer or even advanced cancer, if not diagnosed and treated in time. Liu *et al*. [75] investigated the classification of gastric M-NBI images by fine-turning pre-trained CNNs, which classified them into three classes: chronic gastritis, low grade neoplasia, and EGC. They investigated the performance of four networks: VGG16, InceptionV3, Inception-ResNetV2 and ResNet50, in which ResNet50 got the best result with an accuracy of 0.96.

Esophageal squamous cell carcinoma (ESCC) is one of esophageal cancers. Generally, the basis for the diagnosis of ESCC is histological biopsy. which is a labor intensive task that relies on manual examination and is susceptible to subjective human factors. However, the CAD system based on CNN could provide an objective reference to doctors. Kumagai *et al*. [76] proposed a DL system based on GoogLeNet to identify ESCC from endocytoscopic system (ECS) images of the esophagus to aid confirming histological diagnosis in vivo; the classification accuracy, sensitivity and specificity of this method were 90.9%, 92.6% and 89.3%, respectively. The advantage of this approach is that it can provide objective suggestions for preserving or resecting lesions during examination procedures. There are two limitations of this method. One is that test dataset is too small which may cause low median percentage of the pictures showing malignancy (40.9%) in per-patient analysis. Other limitation is that the images were collected by ESCs with two different optical magnification powers, which may affect the performance. Although these disadvantages, this method still deserves great attention. The invasion depth of ESCC is vital important to the treatment strategy of patients. Nakagawa *et al*. [77] proposed a SSD based system to assess superficial ESCC. This system could classify pathologic mucosal and sub-mucosal micro-invasive (SM1) cancers from submucosal deep invasive (SM2/3) cancers, which is significant for the doctor's choice of patients' treatment strategy.

In all, the applications of DL on the analysis of GI cancer include classification, detection, and recognition tasks, as shown in Table 5, and other tasks such as segmentation are not involved. In other words, more applications of DL on other tasks are encouraged. What's more, the classification of EGC is challenging because some kinds of lesions are difficult to distinguish even for experienced doctors. Even if doctors could identify these cancer lesions, it is still difficult for them to recognize the subtypes. The classification accuracy of EGC is still not satisfactory at present [24], further improvements are still required.

D. MULTIGASTROINTESTINAL DISEASE ANALYSIS

Lesions in GI are diverse, so it is not enough to analysis only a single kind of GI. Recently, some scholars have tried to detect and locate multiple kinds of lesions from GI images.

Since WCE passes through the whole GI, the images collected by WCE are often very large and may contain a variety of lesions. Several scholars have carried out detection of multilesions in WCE images. In [78], Lan *et al*. proposed CNNs based on region proposal algorithm and transfer learning method for the detection of abnormal regions (such as active and inactive bleeding, undigested residue, bubbles, tumor *et al*) in WCE images. The authors also tried several methods and different CNNs. It was indicated that this method was effective for WCE abnormal detection and localization. The advantage of this method is that it could detect and locate multipatterns and multilesions (that is multiobject detection) on a single GI image, which is very different from general single lesion detection or classification focusing on only one disease.

Sekuboyina *et al*. [79] applied CNN models to detect eight different lesions in WCE images, such as bleeding, polyps, ulcers and so on. The authors used a patch-based method. Firstly, the image was divided into several patches; secondly, a CNN was applied to extract features pertaining to

TABLE 5. Overview of papers using dl techniques for gastrointestinal cancer analysis.

Ref.	Application	Remarks	Performances
[24]	EGC classification	Classify M-NBI images into 2 groups: normal and EGC by transfer learning of several popular deep networks	ACC:0.958, SE:0.981, SP:0.989
[68]	GC detection	Using SSD architectures to detect early or AGC	SE:0.922
[69]	Esophageal cancer detection	Using SSD architectures to detect superficial esophageal cancer and advanced cancer	ACC:0.98, SE:0.98
[70]	Esophageal cancer detection	Using transfer learning with CNNs (as feature extractor) to detect squamous cell carcinoma and esophageal adenocarcinoma	AUC:0.92
[71]	GC pathology image recognition	Transfer learning of 4 Deep networks to classify tumors from gastric pathology image	F-score:0.96
[72]	GC images identification	Proposed a DL network called GasricNet to classify GC	ACC:1.00
[73]	GC pathology image classification	Transfer learning by a step-wise manner and using small and big size data to perform classification	-
[74]	GC invasion depth classification	Transfer learning of ResNet50 to determine the invasion depth of GC	ACC:0.8916, SE:74.67%, SP:0.9556
[75]	GC images classification	Transfer learning of 4 popular DL networks to classify gastric images into 3 classes	ACC:0.96
[76]	ESCC classification	Constructed a ESCC classification system based on GoogLeNet	AUC:0.90, SE: 92.60%
[77]	Classification of invasion depth of ESCC	Classify pathologic mucosal, submucosal microinvasive (SM1) cancers and submucosal deep invasive (SM2/3) cancers	SE:90.1%

each patch. This pixel-patch-based framework increased the generality of their method and also overcame the drawbacks caused by artificial features. Similarly, Zhang *et al*. [80] proposed a CNN-based model GPDNet for the classification of three GI diseases: polyps, ulcers, and erosions. The authors introduced an algorithm called iterative reinforced learning (IRL). In this algorithm, the GPDNet was first trained from scratch. Then, a ''fine-tune'' operation through IRL was performed on the model, and the fine-tuned model was used as pre-trained model for further training. The final classification accuracy of this method was 88.9%, which was 8.9% higher than that of the training from scratch. The work of lesion detection and location in a WCE video was also studied by Iakovidis *et al*. [20]. First, using a weakly supervised CNN (WCNN), the authors classified GI endoscopic images into normal and abnormal; second, a deep saliency detection (DSD) algorithm was applied to detect the salient points relevant to these anomalies in endoscopic video frames; third, an iterative cluster unification (ICU) algorithm was applied to locate these anomalies; last, the coordinates of these points were transformed (linearly scaled up) to match the spatial resolution of the input endoscopic image, on which they are superimposed to indicate the possible locations of the anomalies. The detection AUC of this method was 96% and the location AUC was 88%.

Generally, the number of images generated by the WCE is often very large after a WCE examined, and the massive image data analysis may easily result in a misdiagnosis. The GI environment is complex and may be affected by various digestive juices, chyme, bubbles and reflections; as a result, there are a large number of redundant images that could negatively affect diagnosis. It is crucial to filter out these redundant images accurately before classification of the diversity kinds of lesions. However, the deletion of redundant images was not mentioned in the aforementioned literature of this part. In short, the deletion of redundant images need further study attention.

E. OTHER GASTROINTESTINAL DISEASES

Besides these commonly studied GI diseases above, there are some GI diseases with few investigated works, such as gastric ulcer, hookworm infection, Helicobacter pylori (HP) infection, Barrett's esophagus and so on.

Gastric ulcer is one of the common gastric diseases, generally classified as benign and malignant ulcer. Sun *et al*. [67] selected five different CNN models based on VGGNet and IRNV2 (Inception-ResNet) to classify benign and malignant gastric ulcers. The training dataset used in this work contains 854 images with biopsy labels; and the outputted images were marked by a rectangular box with an annotation of the type and a probability score that the lesion belonged to this type (as shown in Fig. 4 (a)). The authors performed several experiments with five models and three data forms, and finally obtained a best classification accuracy rate of 0.866. However, the limited dataset used in this work may have restricted the performance of the method.

Hookworm infection could cause intestinal inflammation and progressive ferritin deficiency anemia, and it can also bring malnutrition and may seriously endanger the health of pregnant women and children. Hookworm detection based on DL was studied by He *et al.* [12]; they designed a method combining two CNNs, one was used in edge extraction, and the other was used in classification. Compared with wu's previous method based on artificial features [13], the accuracy of this method reached 88.5%, which is 10.3% higher than their previous method.

One kind of typical gastritis is caused by HP infection of the gastric mucosa which increases the risk of GC. There are two papers which studied the application of DL in the analysis of HP infection. In [81], the authors carried out transfer learning on a 22-layer pre-trained CNN model to detect HP infected gastritis. Itoh *et al.* [82] also developed a CNN network based on GoogLeNet DCNN pretuned for generic object recognition to analyze HP infection in upper GI. The performances of the two works are shown in Fig. 6.

TABLE 6. Overview of papers using DL techniques for other gastrointestinal diseases analysis.

Ref.	Model	Application	Remarks
[12]	CNN	Hookworm detection	Combined two CNNs, one for edge extraction and one for classification
[67]	VGGNet	Gastric ulcer classification	Classify gastric ulcers into benign and malignant ulcers by a small data
[81]	CNN	HP infection detection	Transfer learning on a 22-layers CNN
[82]	GoogLeNet	HP infection detection	Built a CNN based on GoogLeNet DCNN pretuned generic object recognition system
[83]	CNN	Barrett's esophagus classification	This system could classify two subtype of barrett's esophagus
[84]	SSD	Erosion and ulcerations detection	Trained and validated a new erosion and ulcerations detection system based on SSD
[85]	CNN	Classification of WCE images	GoogLeNet was applied to quantitative analysis and classify celiac patients vs health peoples

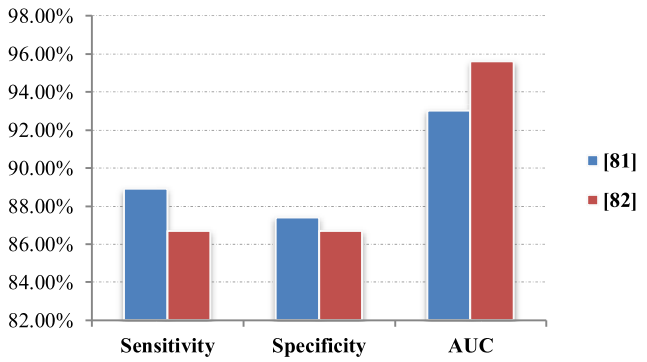


FIGURE 6. Comparison of HP infection detection performances between reference [81] and [82].

We can see that both of them preformed well. However, there is still improvement room for sensitivity and specificity in both of the works.

Barrett's esophagus is a disease that manifests abnormal changes in the cells of esophagus. This disease contains two subtypes, intestinal metaplasia (IM) and gastric metaplasia (GM), in which the IM type could cause neoplasia (NPL) and deteriorate into esophageal cancer. Therefore, it is meaningful to develop DL methods to improve the classification accuracy for IM and GM in the clinic. To solve this problem, the authors of [83] proposed a method based on DL to classify IM, GM and NPL; the classification accuracy of this method reached 80.77% after a data augmentation. However, there is still much improvement room for the accracy.

Detection of erosion and ulcerations in large amount of WCE images is a challenging work. Aoki *et al*. [84] developed a CNN system based on SSD. It was trained by 5560 WCE images of erosion and ulcerations, and validated on a dataset including 10440 WCE images (where only 440 images are abnormal). The processing time of this system only required 233 seconds. The proposed detection system could detect erosion and ulcerations from a large number of WCE images quickly, which could significantly reduce the burden of doctor.

Celiac disease is one of the most common diseases in the world, while few related works are published until now. Zhou *et al*. [85] developed a CAD system based on GoogLeNet model to quantitatively analyze the existence and degree of pathology of the small intestine. This work may improve CAD techniques to access mucosal atrophy and other etiologies in real-time in WCE video.

In summary, there are a few of researches related to these above mentioned diseases, however these diseases pose a great threat to human health. For instance, the HP infection of the gastric mucosa will cause mucosal atrophy and intestinal metaplasia, both of which increase the risk of gastric cancer [82]. Therefore, more researches on these GI diseases are encouraged in the future. Overview of these works is listed in Table 6.

F. OTHER RELATED APPLICATIONS

1) CLASSIFICATION

The classification of WCE images from organ-wise were studied in two papers [86] and [87], which could save the review time of doctors. The authors of [86] proposed a general video understanding approach based on a cascaded spatial-temporal deep framework. The framework mainly consist two CNNs: N-CNN and O-CNN. In the first step, WCE images were classified into informative images and noisy content by the N-CNN model. Then the redundant noisy images were removed. In the second step, the O-CNN was applied to roughly classify the remaining clear images into four digestive organs: entrance, stomach, small intestine and colon. Finally, a hidden Markov model (HMM) coupled with temporal coding observation is applied to further improve the detection accuracy. The system in [87]was designed by combining CNN with extreme learning machine (ELM). The CNN part was used as a data-driven feature extractor and the cascaded ELM as a strong classifier instead of full connected layer. The authors classified the WCE images into three categories: stomach, small intestine and colon. Those approaches could provide organ-wise location information of WCE images to doctors and improve diagnosis efficiency.

Most CAD systems with deep network architectures can only detected very few GI diseases on WCE images. It suggests that the original DL network has to be re-trained when analyzing other GI. For this reason, the authors of [88] introduced a analysis system based on DL methods which learned the generic features of small intestine motion. The advantage of this approach is that it could detect and classify 6 intestinal motility events by one CNN network, and overcoming the problem of re-training network for every new clinical problem.

2) SEGMENTATION

In [89], used an adapted version of SegNet [41] network for reflection region segmentation and color correction. This method may be useful for the preprocessing of GI image before other GI image analysis tasks, such as classification, detection and so on. However, some improvements are still needed to make the reflection correction region smoother.

In all, these applications related to organ-wise classification and reflective elimination which are not analysis GI diseases directly. However, those applications can be used as preprocessing steps for other GI image analysis tasks. For example, the organ-wise classification could provide a organ-wise location for other GI analysis task.

IV. DISCUSSION

In recent years, many efficient deep DL models have emerged with the development of DL theory; they have achieved great success in the field of computer vision and have also been applied to various other fields. Among these DL models, CNNs have performed very well in the field of image processing and has also been applied to analyze various of medical images. Since 2015, DL technique has been gradually utilized to GI image analysis. However, most of the existing research works are still limited to the detection, classification and segmentation of polyps, hemorrhages, GI cancer, and a few works have involved the detection of other diseases, such as esophageal cancer, gastritis and hookworm detection. However, GI diseases are diverse. Some other kind of GI diseases, such as intraepithelial neoplasia and invasive mucosal lesions which are considered to be important stages of early cancer, are also worth to be studied by DL technique, but they have not been mentioned in the existing related literature yet.

The application of the DL technique in computer-aided GI diagnosis is a new research field. In this review, several key words, such as gastrointestinal, deep learning, CNN, digestive tract and lesion detection, were used to retrieve the latest relevant literature. In all, 45 papers were found, most of which were published during 2017-2019; these papers consisted of 11 papers related to polyp, 8 papers related to intestinal hemorrhage, 11 papers related to GI cancer, 4 papers about multi-GI diseases, 7 papers related other GI diseases that are not commonly studied such as HP infection, hookworm detection, and 4 papers are related to other applications. The statistic of these papers are shown in Fig. 7. We can see from Fig. 7 (a) that more than half of these published researches focus on the detection and classification tasks. Other analysis tasks such as segmentation and recognition are only studied by few papers, so in the future these tasks deserved further study. Fig. 7 (b) presents the proportion of each GI disease that was included in the related literature. The polyp is the most popular studied GI disease, and papers proportion about GI cancer are the second most popular studied. The rest GI diseases are not commonly studied such as HP infection and hookworm detection, which keep promising researches in the future. Fig. 7 (c) shows the proportion of papers counted according to three endoscopies. We can see that WCE is the most popular. Fig. 7 (d), provides a statistical data of the number of related published papers vs year. The earliest research on the application of DL methods in GI image appeared around 2015 and the literature number keeps increasing with year. Up to now, there have been published 11 papers related to GI diseases this year, and according to this trend more papers will be published in 2019. Studies [21], [23], [25], [52], and [67] are ground-breaking works on the application of DL techniques in GI diseases.

DL method requires a large number of labeled training data sets. For example, the training dataset of AlexNet contains 1.2 million samples. Due to the high cost of manual labeling by medical experts and the consideration of patient privacy issues, it is difficult to obtain a large amount of labeled medical image data. Unlike to skin images, eye images, MR and CT images which are collected from the body surface, the collection of GI image requires performing an endoscopy, which involves entering a camera probe into the patient's body. Therefore, the data acquisition of GI image is more difficult, and the application of DL in computer-aided GI diagnosis is severely limited, challenging and nonproductive. Great success has been achieved in the application of DL in other diseases, such as eye [19] and skin diseases [18], owing to the sufficient training data sets that contain more than 100,000 labeled images.

Moreover, detected objects in natural images are often colorful with clear boundaries, while lesions found in medical images lack a standardized, consistent shape, and do not always have clear edges. Considering these differences between the natural and medical images, models trained on natural images may not be useful to assess medical images well. Moreover, if the training data set is insufficient during the transfer learning process, the resulting analysis may be unremarkable. The differences among several common types of medical images are smaller than those of natural images; if transfer learning is performed on a foundation of the models that are pre-trained with medical images, the results may be better than those of directly using natural image pre-trained models.

Insufficient image data and unbalanced samples are common problems faced by the application of DL in all kinds of medical image analysis. Data augmentation [90] could overcome this problem effectively and help reducing overfitting, and could also improve the stability and classification accuracy at the same time [21]. Conventional data augmentation methods generally include rotation, flipping, shading, scaling, and affine transformation. Until now, there has been no DL model that is completely trained on huge medical image data (the data scale as ImageNet) from scratch. In all, the application of DL in the field of medical image processing remains immature but still has great potential. Whether a network model is good or not is affected by many factors, such as image quality, sharpness, and label accuracy that can affect training results. The lack of a common validation framework is a major problem in medical and endoscopic image analysis [91], which limits the effectiveness of comparisons

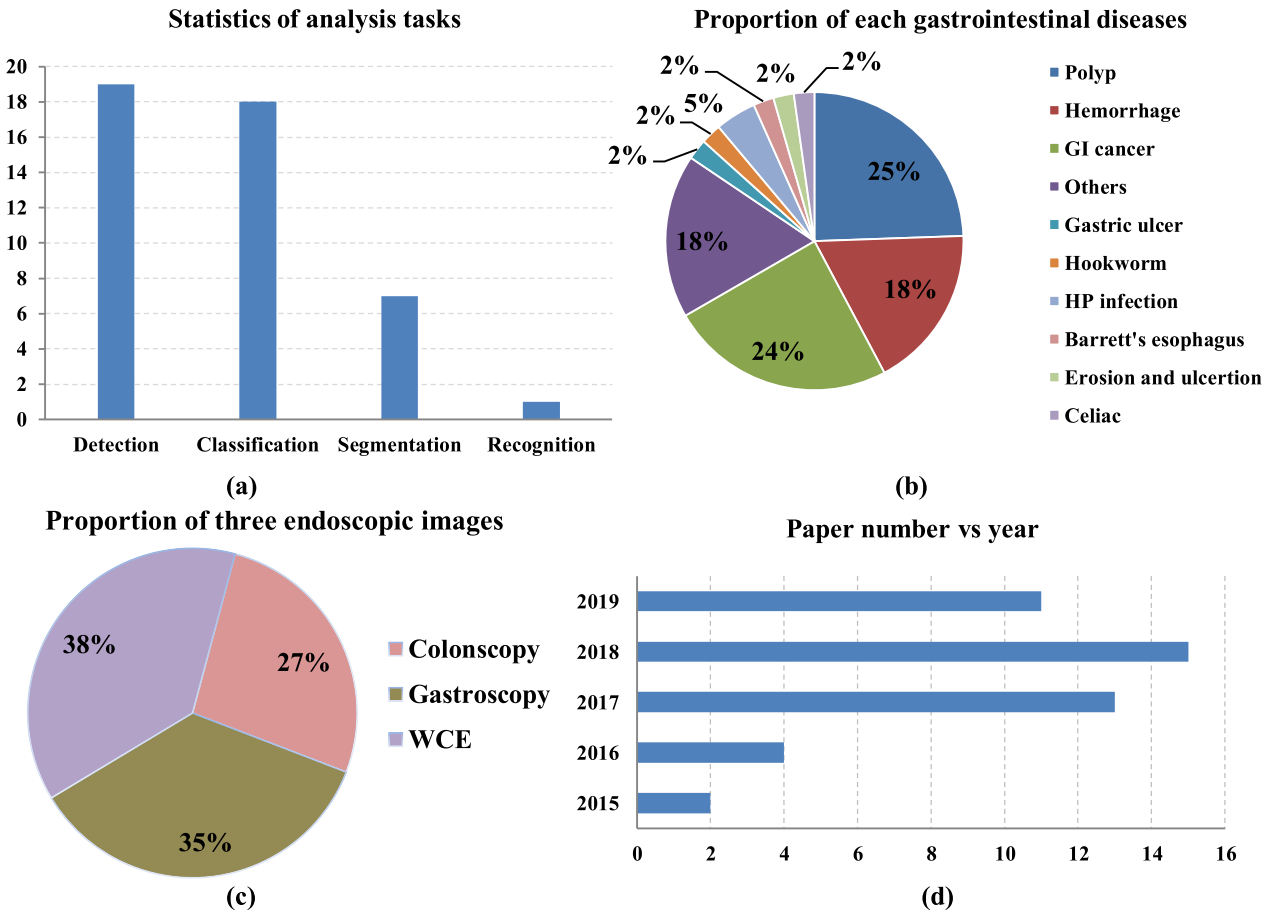


FIGURE 7. The statistic data of papers related to GI diseases cited in this review. (a) The paper number of different analysis tasks; (b) the percentage of published papers relevant to different GI diseases; (c) the proportion of papers on three different endoscopic image; (d) The number of paper published each year.

between existing methods and makes it difficult to conclude which one contributes more to the practical and clinical application. We hope that a large public image database (similar to ImageNet) containing all kinds of medical images with considerable amount of data can be built in the future, which could provide enough data to researchers and further promote the application of DL techniques in medical image analysis. Thus, a new revolution in AI-based medical diagnosis could be achieved.

DL methods can be divided into supervised learning and unsupervised learning methods. Currently, almost all of the deep models used in GI image processing are CNN-based supervised learning networks, only one was based on GANs segmentation [64]. Supervised learning requires labeled training data, but the production of labels is subjective and costly. There may be great visual differences among some images of the same disease, and there may also be slight differences among images of different diseases. As a result, different endoscopic experts may give different labels to the same image [92]. These controversial and incorrectly labeled data may mislead the network and slow down the training process. However, in unsupervised learning, which only uses unlabeled data for training, the above problem does not exist. An unsupervised deep network can detect subtle features that can barely be detected by human eyes, so this kind of DL methods is competent enough to classify these controversial images into the correct categories. Deep networks that are trained in an unsupervised manner may be more adaptive to a dataset with poor labeling accuracy. In short, unsupervised DL deserves further exploration in the field of GI image processing.

V. CONCLUSION

GI image analysis is a new application field of DL methods. It has not been widely applied in GI images analysis until the last few years that a small group of scholars have tried to study in this field.

Although some results have been achieved, the researches related to the application of DL in this field is still rare, and the potential of this technique is far from being fully explored. Several aspects of DL-based GI image analysis deserve further study: (1) Development of a 3D-CNN based DL diagnostic system. 3D-CNN can learn more spatial features and better encode the spatial information; (2) Development of a real-time detection system. Many GI surgeries are endoscopic surgeries. If real-time endoscopic diagnosis

TABLE 7. Summary of acronyms or abbreviations and the corresponding full names.

Full name	Acronyms
Advanced gastric cancer	AGC
Artificial intelligence	AI
Artificial neural network	ANN
Area under curve	AUC
Atrous spatial pyramid pooling	ASPP
Computer aided dignosis	CAD
Canonical correlation analysis networks	CCANet
Conditional random field	CRF
Convolutional neural network	CNN
Deep convolutional neural network	DCNN
Deep learning	DL
Deep neural network	DNN
Deep saliency detection	DSD
Early gastric cancer	EGC
Endocytoscopic system	ECS
Esophageal squamous cell carcinoma	ESCC
Extreme learning machine	ELM
False positive	FP
Fully convolutional network	FCN
Gastric cancer	GC
Gastric metaplasia	GM
Gastrointestinal	GI
Generative adversarial network	GAN
Graph neural network	GNN
Hidden Markov model	HMM
ImageNet large-scale visual recognition challenge	ILSVRC
Intestinal meteplasia	IM
Iterative cluster unification	ICU
Iterative reinforced learning	IRL
Long short-term memory	LSTM
Machine learning	ML
Magnification endoscopy with Narrow-band imaging	M-NBI
Mean intersection over union	mIOU
Medical image computing and computer assisted intervention	MICCAI
Narrow-band imaging	NBI
Neoplasia	NPL
Principle component analysis network	PCANet
Receiver operating characteristic	ROC
Rectified linear unit	ReLU
Recurrent neural network	RNN
Region-based convolutional network	R-CNN
Single shot multi-box detection	SSD
Support vector machine	SVM
Weakly supervised CNN	WCNN
Wireless capsule endoscopy	WCE

system can be used with high-precision and efficiency, surgeries could be directly performed during the examination, and the histopathological biopsy step would no longer be indispensable. Thus, the suffering of patients could be reduced; (3) Improvement of the detection accuracy of early cancer. As the five-year survival rate of EGC is as high as 95%, increasing the detection accuracy and reducing the false negatives rate of cancer in the early stage are critical for making early treatment available to every early cancer patient; (4) Development of a diagnostic system based on an unsupervised learning method. Unsupervised DL diagnostic systems can alleviate the problem brought by ''no label'' or ''controversial labels''. (5) Assessment of the invasion depth of cancers, which is utmost important for the treat strategy of cancer patients. (6) Development of other DL methods such as RNNs and GNNs, which hold promising in GI image analysis. In short, DL is with great potential and may play an important role in the clinical aided-diagnosis of GI in the future.

APPENDIX

Because there are so many acronyms or abbreviations in this paper, here we summarize them with the corresponding full name in Table 7.

REFERENCES

[1] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, ''Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries,'' *CA, Cancer J. Clin.*, vol. 68, no. 6, pp. 394–424, 2018.

[2] A. Gondosa, F. Bray, D. H. Brewster, J. W. W. Coebergh, M. L. G. Janssen-Heijnen, J. Kurtinaitis, H. Brenner, T. Hakulinen, and The EUNICE Survival Working Group, ''Recent trends in cancer survival across Europe between 2000 and 2004: A model-based period analysis from 12 cancer registries,'' *Eur. J. Cancer*, vol. 44, no. 10, pp. 1463–1475, Jul. 2008.

[3] K. Washington, ''7th edition of the AJCC cancer staging manual: Stomach,'' *Ann. Surgical Oncol.*, vol. 17, no. 12, pp. 3077–3079, 2010.

[4] I. Take, Q. Shi, and Y.-S. Zhong, ''Progress with each passing day: Role of endoscopy in early gastric cancer,'' *Transl. Gastrointest. Cancer*, vol. 4, no. 6, pp. 423–428, 2015.

[5] J. Mannath and K. Ragunath, ''Role of endoscopy in early oesophageal cancer,'' *Nature Rev. Gastroenterol. Hepatol.*, vol. 13, no. 12, pp. 720–730, 2016.

[6] D. Liu, N. Rao, X. Mei, H. Jiang, Q. Li, C. Luo, Q. Li, C. Zeng, B. Zeng, and T. Gan, ''Annotating early esophageal cancers based on two saliency levels of gastroscopic images,'' *J. Med. Syst.*, vol. 42, no. 12, p. 237, 2018.

[7] D.-Y. Liu, G. Tao, N.-N. Rao, X. Yao-Wen, Z. Jie, L. Sang, L. Cheng-Si, Z. Zhong-Jun, and W. Yong-Li, ''Identification of lesion images from gastrointestinal endoscope based on feature extraction of combinational methods with and without learning process,'' *Med. Image Anal.*, vol. 32, pp. 281–294, Aug. 2016.

[8] D. Y. Liu, N. N. Rao, X. M. Mei, C. S. Luo, Y. W. Xing, and T. Gan, ''An automatic annotation method for early esophageal cancers based on saliency guided superpixel segmentation,'' in *Proc. ICBCI*, Beijing, China, Sep. 2017, pp. 43–47.

[9] D. Y. Liu, T. Gan, N. N. Rao, G. G. Xu, B. Zeng, and H. L. Li, ''Automatic detection of early gastrointestinal cancer lesions based on optimal feature extraction from gastroscopic images,'' *J. Med. Imag. Heal. Inf.*, vol. 5, no. 2, pp. 296–302, 2015.

[10] Y. Shin and I. Balasingham, ''Comparison of hand-craft feature based SVM and CNN based deep learning framework for automatic polyp classification,'' in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBS)*, Jeju Island, South Korea, Jul. 2017, pp. 3277–3280.

[11] 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.

[12] J.-Y. He, X. Wu, Y.-G. Jiang, Q. Peng, and R. Jain, ''Hookworm detection in wireless capsule endoscopy images with deep learning,'' *IEEE Trans. Image Process.*, vol. 27, no. 5, pp. 2379–2392, May 2018.

[13] X. Wu, H. Chen, T. Gan, J. Chen, C.-W. Ngo, and Q. Peng, ''Automatic hookworm detection in wireless capsule endoscopy images,'' *IEEE Trans. Med. Imag.*, vol. 35, no. 7, pp. 1741–1752, Jul. 2016.

[14] A. Krizhevsky, I. Sutskever, and G. E. Hinton, ''ImageNet classification with deep convolutional neural networks,'' in *Proc. Adv. Neural Inf. Process. Syst.*, Dec. 2012, pp. 1097–1105.

[15] K. Fukushima, ''Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position,'' *Biol. Cybern.*, vol. 36, no. 4, pp. 193–202, Apr. 1980.

[16] S. C. B. Lo, S. L. A. Lou, J.-S. Lin, M. T. Freedman, M. V. Chien, and S. K. Mun, ''Artificial convolution neural network techniques and applications for lung nodule detection,'' *IEEE Trans. Med. Imag.*, vol. 14, no. 4, pp. 711–718, Dec. 1995.

[17] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

[18] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, 2017.

[19] V. Gulshan, L. Peng, M. Coram, M. C. Stumpe, D. Wu, A. Narayanaswamy, S. Venugopalan, K. Widner, T. Madams, J. Cuadros, and R. Kim, "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," *J. Amer. Med. Assoc.*, vol. 316, no. 22, pp. 2402–2410, 2016.

[20] D. K. Iakovidis, S. V. Georgakopoulos, M. Vasilakakis, A. Koulaouzidis, and V. P. Plagianakos, "Detecting and locating gastrointestinal anomalies using deep learning and iterative cluster unification," *IEEE Trans. Med. Imag.*, vol. 37, no. 10, pp. 2196–2210, Oct. 2018.

[21] Y. Shin, H. A. Qadir, L. Aabakken, J. Bergsland, and I. Balasingham, "Automatic colon polyp detection using region based deep cnn and post learning approaches," *IEEE Access*, vol. 6, pp. 40950–40962, 2018.

[22] M. F. Byrne, "Real-time differentiation of adenomatous and hyperplastic diminutive colorectal polyps during analysis of unaltered videos of standard colonoscopy using a deep learning model," *Gut*, vol. 68, no. 1, pp. 94–100, 2019.

[23] X. Jia and M. Q.-H. Meng, "A deep convolutional neural network for bleeding detection in wireless capsule endoscopy images," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBS)*, Kuala, Lumpur, Aug. 2016, pp. 639–642.

[24] X. Liu, C. Wang, Y. Hu, Z. Zeng, J. Bai, and G. Liao, "Transfer learning with convolutional neural network for early gastric cancer classification on magnifiying narrow-band imaging images," in *Proc. Int. Conf. Image Process. (ICIP)*, Athens, Greece, Oct. 2018, pp. 1388–1392.

[25] L. Yu, H. Chen, Q. Dou, J. Qin, and P. A. Heng, "Integrating online and offline three-dimensional deep learning for automated polyp detection in colonoscopy videos," *IEEE J. Biomed. Health Inform.*, vol. 21, no. 1, pp. 65–75, Jan. 2017.

[26] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. W. M. van der Laak, B. van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, Dec. 2017.

[27] J. Ker, L. Wang, J. Rao, and T. Lim, "Deep learning applications in medical image analysis," *IEEE Access*, vol. 6, pp. 9375–9389, 2018.

[28] D. Shen, G. Wu, and H. Suk, "Deep learning in medical image analysis," *Annu. Rev. Biomed. Eng.*, vol. 19, pp. 221–248, Jun. 2017.

[29] M. K. Chandrakar and A. Mishra, "Review of medical image analysis, segmentation and application using deep learning," *J. Adv. Res. Dyn. Control Syst.*, vol. 10, no. 1, pp. 549–553, 2018.

[30] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015.

[31] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. ICLR*, San, Diego, CA, USA, 2015.

[32] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Boston, MA, USA, Jun. 2015, pp. 1–9.

[33] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, inception-ResNet and the impact of residual connections on learning," Aug. 2016, *arXiv:1602.07261*. [Online]. Available: https://arxiv.org/abs/1602.07261

[34] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Las Vegas, NV, USA, Jun. 2016, pp. 770–778.

[35] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Salt Lake, UT, USA, Jun. 2018, pp. 7132–7141.

[36] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single shot multibox detector," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Amsterdam, The Netherlands, 2016, pp. 21–37.

[37] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis.*, Santiago, Chile, Dec. 2015, pp. 1440–1448.

[38] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.

[39] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 640–651, Apr. 2017.

[40] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs.," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Apr. 2018.

[41] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.

[42] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolu-tional networks for biomedical image segmentation," in *Proc. Med. Image Comput. Comput.-Assist. Intervent.*, Munich, Germany, 2016, pp. 234–241.

[43] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, Kuching, Malaysia, 2014, pp. 1–9.

[44] F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner, and G. Monfardini, "The graph neural network model," *IEEE Trans. Neural Netw.*, vol. 20, no. 1, pp. 61–80, Jan. 2009.

[45] T.-H. Chan, K. Jia, S. Gao, J. Lu, Z. Zeng, and Y. Ma, "PCANet: A simple deep learning baseline for image classification?" *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 5017–5032, Dec. 2015.

[46] X. Yang, W. Liu, D. Tao, and J. Cheng, "Canonical correlation analysis networks for two-view image recognition," *Inf. Sci.*, vols. 385–386, pp. 338–352, Apr. 2017.

[47] S. Andermatt, S. Pezold, and P. Cattin, "Multi-dimensional gated recurrent units for the segmentation of biomedical 3D-data," in *Deep Learning and Data Labeling for Medical Applications*. Athens, Greece, 2016, pp. 142–151.

[48] Y. Bengio, P. Simard, and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," *IEEE Trans. Neural Netw.*, vol. 5, no. 2, pp. 157–166, Mar. 1994.

[49] X. Gao, S. Lin, and T. Y. Wong, "Automatic feature learning to grade nuclear cataracts based on deep learning," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 11, pp. 2693–2701, Nov. 2015.

[50] J. Zhou, G. Cui, Z. Zhang, C. Yang, Z. Liu, L. Wang, C. Li, and M. Sun, "Graph neural networks: A review of methods and applications," 2018, *arXiv:1812.08434*. [Online]. Available: https://arxiv.org/abs/1812.08434

[51] 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.

[52] R. Zhang, Y. Zheng, T. W. C. Mak, R. Yu, S. H. Wong, J. Y. W. Lau, and C. C. Y. Poon, "Automatic detection and classification of colorectal polyps by transferring low-level CNN features from nonmedical domain," *IEEE J. Biomed. Health Inform.*, vol. 21, no. 1, pp. 41–47, Jan. 2017.

[53] E. Ribeiro, A. Uhl, and M. Häfner, "Colonic polyp classification with convolutional neural networks," in *Proc. IEEE 29th Int. Symp. Comput.-Based Med. Syst.*, Dublin, Ireland, Jun. 2016, pp. 253–258.

[54] W. E. Karnes, T. Alkayali, M. Mittal, A. Patel, J. Kim, K. J. Chang, A. Q. Ninh, G. Urban, and P. Baldi, "Su1642 automated polyp detection using deep learning: Leveling the field," *Gastrointestinal Endoscopy*, vol. 85, no. 5, pp. AB376–AB377, 2017.

[55] E. Ribeiro, A. Uhl, G. Wimmer, and M. Häfner, "Exploring deep learning and transfer learning for colonic polyp classification," *Comput. Math. Methods Med.*, vol. 2016, Oct. 2016, Art. no. 6584725.

[56] P.-J. Chen, M.-C. Lin, M.-J. Lai, J.-C. Lin, H. H.-S. Lu, and V. S. Tseng, "Accurate classification of diminutive colorectal polyps using computer-aided analysis," *Gastroenterology*, vol. 154, no. 3, pp. 568–575, 2018.

[57] N. Tajbakhsh, S. R. Gurudu, and J. Liang, "Automatic polyp detection in colonoscopy videos using an ensemble of convolutional neural networks," in *Proc. Int. Symp. Biomed. Imag.*, New York, NY, USA, Apr. 2015, pp. 79–83.

[58] W.-T. Xiao, L.-J. Chang, and W.-M. Liu, "Semantic segmentation of colorectal polyps with DeepLab and LSTM networks," in *Proc. IEEE Int. Conf. Consum. Electron.-Taiwan*, Taiwan, China, May 2018, pp. 1–2.

[59] P. Li, Z. Li, F. Gao, L. Wan, and J. Yu, "Convolutional neural networks for intestinal hemorrhage detection in wireless capsule endoscopy images," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Hong Kang, China, Jul. 2017, pp. 1518–1523.

[60] X. Jia and M. Q.-H. Meng, "Gastrointestinal bleeding detection in wireless capsule endoscopy images using handcrafted and CNN features," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBS)*, Jeju-do, South Korea, Jul. 2017, pp. 3154–3157.

[61] S. Tanabe, "Diagnosis of obscure gastrointestinal bleeding," *Clin. Endosc.*, vol. 49, no. 6, pp. 539–541, 2016.

[62] X. Jia and M. Q.-H. Meng, "A study on automated segmentation of blood regions in wireless capsule endoscopy images using fully convolutional networks," in *Proc. Int. Symp. Biomed. Imag.*, Melbourne, VIC, Australia, Apr. 2017, pp. 179–182.

[63] R. Leenhardt, P. Vasseur, C. Li, J. C. Saurin, G. Rahmi, F. Cholet, A. Becq, P. Marteau, A. Histace, and X. Dray, "A neural network algorithm for detection of GI angiectasia during small-bowel capsule endoscopy," *Gastrointestinal Endoscopy*, vol. 89, no. 1, pp. 189–194, Jan. 2019.

[64] K. Pogorelov, O. Ostroukhova, A. Petlund, P. Halvorsen, T. de Lange, H. N. Espeland, T. Kupka, C. Griwodz, and M. Riegler, "Deep learning and handcrafted feature based approaches for automatic detection of angiectasia," in *Proc. IEEE EMBS Int. Conf. Biomed. Health Inform. (BHI)*, Las Vegas, NV, USA, Mar. 2018, pp. 365–368.

[65] T. Ghosh, L. Li, and J. Chakareski, "Effective deep learning for semantic segmentation based bleeding zone detection in capsule endoscopy images," in *Proc. Int. Conf. Image Process. (ICIP)*, Athens, Greece, Oct. 2018, pp. 3034–3038.

[66] M. Hajabdollahi, R. Esfandiarpoor, P. Khadivi, S. M. R. Soroushmehr, N. Karimi, K. Najarian, and S. Samavi, "Segmentation of bleeding regions in wireless capsule endoscopy for detection of informative frames," *Biomed. Signal Process. Control*, vol. 53, Aug. 2019, Art. no. 101565.

[67] J. Y. Sun, S. W. Lee, M. C. Kang, S. W. Kim, S. Y. Kim, and S. J. Ko, "A novel gastric ulcer differentiation system using convolutional neural networks," in *Proc. IEEE Symp. Comput.-Based Med. Syst.*, Karlstad, Sweden, Jun. 2018, pp. 351–356.

[68] T. Hirasawa, K. Aoyama, T. Tanimoto, S. Ishihara, S. Shichijo, T. Ozawa, T. Ohnishi, M. Fujishiro, K. Matsuo, J. Fujisaki, and T. Tada, "Application of artificial intelligence using a convolutional neural network for detecting gastric cancer in endoscopic images," *Gastric Cancer*, vol. 21, no. 4, pp. 653–660, 2018.

[69] Y. Horie, T. Yoshio, K. Aoyama, S. Yoshimizu, Y. Horiuchi, A. Ishiyama, T. Hirasawa, T. Tsuchida, T. Ozawa, S. Ishihara, Y. Kumagai, M. Fujishiro, I. Maetani, J. Fujisaki, and T. Tada, "Diagnostic outcomes of esophageal cancer by artificial intelligence using convolutional neural networks," *Gastrointestinal Endoscopy*, vol. 89, no. 1, pp. 25–32, 2019.

[70] S. Van Riel, F. Van Der Sommen, S. Zinger, E. J. Schoon, and P. H. N. De With, "Automatic detection of early esophageal cancer with CNNS using transfer learning," in *Proc. Int. Conf. Image Process. (ICIP)*, Athens, Greece, Oct. 2018, pp. 1383–1387.

[71] B. Liu, K. Yao, M. Huang, J. Zhang, Y. Li, and R. Li, "Gastric pathology image recognition based on deep residual networks," in *Proc. Int. Comput. Softw. Appl. Conf.*, Tokyo, Japan, Jul. 2018, pp. 408–412.

[72] Y. Li, X. Li, X. Xie, and L. Shen, "Deep learning based gastric cancer identification," in *Proc. Int. Symp. Biomed. Imag.*, Washington DC, USA, Apr. 2018, pp. 182–185.

[73] J. Qu, N. Hiruta, K. Terai, H. Nosato, M. Murakawa, and H. Sakanashi, "Gastric pathology image classification using stepwise fine-tuning for deep neural networks," *J. Healthcare Eng.*, vol. 2018, Jun. 2018, Art. no. 8961781.

[74] Y. Zhu, Q.-C. Wang, M.-D. Xu, Z. Zhang, J. Cheng, Y.-S. Zhong, Y.-Q. Zhang, W.-F. Chen, L.-Q. Yao, P.-H. Zhou, and Q.-L. Li, "Application of convolutional neural network in the diagnosis of the invasion depth of gastric cancer based on conventional endoscopy," *Gastrointestinal Endoscopy*, vol. 89, no. 4, pp. 806–815, 2019.

[75] X. Liu, C. Wang, J. Bai, and G. Liao, "Fine-tuning pre-trained convolutional neural networks for gastric precancerous disease classification on magnification narrow-band imaging images," *Neurocomputing*, to be published. doi: 10.1016/j.neucom.2018.10.100.

[76] Y. Kumagai, K. Takubo, K. Kawada, K. Aoyama, Y. Endo, T. Ozawa, T. Hirasawa, T. Yoshio, S. Ishihara, M. Fujishiro, J.-I. Tamaru, E. Mochiki, H. Ishida, and T. Tada, "Diagnosis using deep-learning artificial intelligence based on the endocytoscopic observation of the esophagus," *Esophagus*, vol. 16, no. 2, pp. 180–187, 2019.

[77] K. Nakagawa, R. Ishihara, K. Aoyama, K. Aoyama, M. Ohmori, H. Nakahira, N. Matsuura, S. Shichijo, T. Nishida, T. Yamada, S. Yamaguchi, H. Ogiyama, S. Egawa, O. Kishida, and T. Tada, "Classification for invasion depth of esophageal squamous cell carcinoma using a deep neural network compared with experienced endoscopists," *Gastrointestinal Endoscopy*, vol. 90, no. 3, pp. 407–414, Sep. 2019. doi: 10.1016/j.gie.2019.04.245.

[78] L. Lan, C. Ye, C. Wang, and S. Zhou, "Deep convolutional neural networks for WCE abnormality detection: CNN architecture, region proposal and transfer learning," *IEEE Access*, vol. 7, pp. 30017–30032, 2019.

[79] A. K. Sekuboyina, S. T. Devarakonda, and C. S. Seelamantula, "A convolutional neural network approach for abnormality detection in wireless capsule endoscopy," in *Proc. Int. Symp. Biomed. Imag.*, Melbourne, VIC, Australia, Apr. 2017, pp. 1057–1060.

[80] X. Zhang, W. Hu, F. Chen, J. Liu, Y. Yang, L. Wang, H. Duan, and J. Si, "Gastric precancerous diseases classification using CNN with a concise model," *PLoS ONE*, vol. 12, no. 9, 2017, Art. no. e0185508.

[81] S. Shichijo, S. Nomura, K. Aoyama, Y. Nishikawa, M. Miura, T. Shinagawa, H. Takiyama, T. Tanimoto, S. Ishihara, K. Matsuo, and T. Tada, "Application of convolutional neural networks in the diagnosis of *helicobacter pylori* infection based on endoscopic images," *EBioMedicine*, vol. 25, pp. 106–111, Nov. 2017.

[82] T. Itoh, H. Kawahira, H. Nakashima, and N. Yata, "Deep learning analyzes Helicobacter pylori infection by upper gastrointestinal endoscopy images," *Endosc. Int. Open*, vol. 6, no. 2, pp. E139–E144, 2018.

[83] J. Hong, B.-Y. Park, and H. Park, "Convolutional neural network classifier for distinguishing Barrett's esophagus and neoplasia endomicroscopy images," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBS)*, Seogwipo, South Korea, Jul. 2017, pp. 2892–2895.

[84] T. Aoki, A. Yamada, K. A. MMath, H. Saito, A. Tsuboi, A. Nakada, R. Niikura, M. Fujishiro, S. Oka, S. Ishihara, T. Matsuda, S. Tanaka, K. Koike, and T. Tada, "Automatic detection of erosions and ulcerations in wireless capsule endoscopy images based on a deep convolutional neural network," *Gastrointestinal Endoscopy*, vol. 89, no. 2, pp. 357–363, 2019.

[85] T. Zhou, G. Han, B. N. Li, Z. Lin, E. J. Ciaccio, P. H. Green, and J. Qin, "Quantitative analysis of patients with celiac disease by video capsule endoscopy: A deep learning method," *Comput. Biol. Med.*, vol. 85, pp. 1–6, Jun. 2017.

[86] H. Chen, X. Wu, G. Tao, and Q. Peng, "Automatic content understanding with cascaded spatial–temporal deep framework for capsule endoscopy videos," *Neurocomputing*, vol. 229, Mar. 2017, pp. 77–87.

[87] J.-S. Yu, J. Chen, Z. Q. Xiang, and Y.-X. Zou, "A hybrid convolutional neural networks with extreme learning machine for WCE image classification," in *Proc. IEEE Int. Conf. Robot. Biomimetics*, Zhuhai, China, Dec. 2015, pp. 1822–1827.

[88] S. Seguí, M. Drozdzal, G. Pascual, P. Radeva, C. Malagelada, F. Azpiroz, and J. Vitrià, "Generic feature learning for wireless capsule endoscopy analysis," *Comput. Biol. Med.*, vol. 79, pp. 163–172, Dec. 2016.

[89] A. Rodríguez-Sánchez, D. Chea, G. Azzopardi, and S. Stabinger, "A deep learning approach for detecting and correcting highlights in endoscopic images," in *Proc. 7th Int. Conf. Image Process. Theory, Tools Appl. (IPTA)*, Montreal, QC, Canada, Nov./Dec. 2017, pp. 1–6.

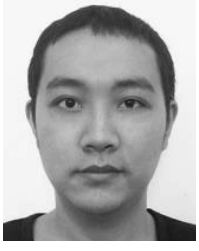
[90] A. Asperti and C. Mastronardo, "The effectiveness of data augmentation for detection of gastrointestinal diseases from endoscopical images," Dec. 2017, *arXiv:1712.03689*. [Online]. Available: https://arxiv.org/abs/1712.03689

[91] D. K. Iakovidis and A. Koulaouzidis, "Software for enhanced video capsule endoscopy: Challenges for essential progress," *Nature Rev. Gastroenterol. Hepatol.*, vol. 12, no. 3, pp. 172–186, 2015.

[92] M. J. J. P. van Grinsven, B. van Ginneken, C. B. Hoyng, T. Theelen, and C. I. Sánchez, "Fast convolutional neural network training using selective data sampling: Application to hemorrhage detection in color fundus images," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1273–1284, May 2016.

WENJU DU was born in Shandong, China. She received the B.Eng. degree in biomedical engineering, in 2014, and the master's degree from the School of Life Science and Technology, University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2015, where she is currently pursuing the master's and Ph.D. combined degree in biomedical engineering. Her research interests include the application of deep learning on gastrointestinal image analysis, and fuzzy control and switched systems.

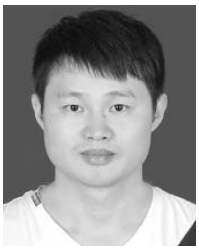
NINI RAO received the B.S. and M.S. degrees in electronic engineering and the Ph.D. degree in biomedical engineering from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 1983, 1989, and 2009, respectively. She was a Visiting Scholar with the University of Georgia and with the Massachusetts General Hospital/Harvard Medical School, from April 2008 to October 2008 and from March 2016 to September 2016, respectively, and a Visiting Professor with the National University of Singapore, from July 2006 to August 2006. She is currently a Professor with the School of Life Science and Technology, UESTC. She is the author or coauthor of more than 150 scientific articles. Her major research interests include biomedical signal and image processing, biomedical pattern recognition, and bioinformatics. She was a recipient of more than 20 research grants. She was honored with the Outstanding Expert with Outstanding Contribution to Sichuan province, in 2005, and the Academic and Technical Leader in Sichuan province, in 2011. She was also a recipient of the Third-Class Prize of progress of science and technology of Sichuan Province, in 2012.



DINGYUN LIU was born in Beijing, China, in 1990. He received the bachelor's degree in electronic engineering from the School of electronic engineering, University of Electronic Science and Technology of China (UESTC), and the master's degree in biomedical engineering from the School of Life Science and Technology, UESTC, in 2012, where he is currently pursuing the master's and Ph.D. combined degree. He has published one patent and more than ten research articles on the above research field. His research interests include gastrointestinal endoscopic image processing, such as the detection and annotation of early gastric cancer, esophageal cancer, and the abnormal frame detection in wireless capsule endoscopic images, ECG signal processing, such as the detection of atrial fibrillation, and bio-informatics of gastric cancer.

HONGXIU JIANG received the B.Eng. degree in biomedical engineering from the Southwest University of Science and Technology, in 2017. She is currently pursuing the M.S. degree with the School of Life Science and Technology, University of Electronic Science and Technology of China, Chengdu, China. Her research interest includes medical image/video processing.

CHENGSI LUO received the B.Sc. and M.Sc. degrees from the School of Biomedical Engineering, University of Electronic Science and Technology of China, in 2015 and 2018, respectively, where he is currently pursuing the Ph.D. degree. His research interests include biomedical signal processing and deep convolutional neural networks.



ZHENGWEN LI was born in Shandong, China, in 1984. He received the master's degree in applied mathematics from Southwest Jiaotong University, in 2010. He is currently pursuing the Ph.D. degree with the College of Life Sciences, University of Electronic Science and Technology of China.

He was a Math Teacher with the Chengdu College of Electronic Science and Technology University, for four years. Since 2016, he has mainly studied the application of convolutional neural networks in gastroscopic images and published several related articles.

TAO GAN received the master's degree in medical in digestion from the West China College, Sichuan, China, in 2000, and the master's degree in computer science from the University of Electronic Science and Technology of China, Sichuan, in 2006.

He is currently an Associate Professor with West China Hospital, Sichuan University. He has been an in charge of provincial scientific project. He has published over than 20 articles. He holds three Chinese patents. His research interests include medical image processing and digestive endoscopy.

BING ZENG (M'91–SM'13–F'16) received the B.Eng. and M.Eng. degrees in electronic engineering from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 1983 and 1986, respectively, and the Ph.D. degree in electrical engineering from the Tampere University of Technology, Tampere, Finland, in 1991.

He was a Postdoctoral Fellow with the University of Toronto, from September 1991 to July 1992, and a Researcher with Concordia University, from August 1992 to January 1993. He joined the Hong Kong University of Science and Technology (HKUST). After 20 years of service at HKUST, he returned to UESTC, in summer of 2013, through China's 1000-Talent-Scheme. At UESTC, he leads the Institute of Image Processing to work on image and video processing, 3D and multiview video technology, and visual big data. During his tenure at HKUST and UESTC, he graduated more than 40 Master and Ph.D. students. He was a recipient of about 20 research grants and filed eight international patents. He has published more than 260 articles. Three representing works are as follows: one article on fast block motion estimation, published in the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY (TCSVT), in 1994. He was elected as an IEEE Fellow in 2016 for contributions to image and video coding. He has been SCI-cited more than 1000 times (Google-cited more than 2200 times) and currently stands at the 8th position among all articles published in this Transactions; one article on smart padding for arbitrarily-shaped image blocks, published in the IEEE TCSVT, in 2001, led to a patent that has been successfully licensed to companies; and one article on directional discrete cosine transform (DDCT), published in the IEEE TCSVT, in 2008,received the 2011 IEEE CSVT Transactions Best Paper Award. He was also a recipient of the Best Paper Award at ChinaCom three times (2009 Xi'an, 2010 Beijing, and 2012 Kunming), the Best Associate Editor Award, in 2011, and the 2nd Class Natural Science Award (the first recipient) from the Chinese Ministry of Education, in 2014. He served as an Associate Editor for the IEEE TCSVT for eight years. He was the General Co-Chair of VCIP-2016 and PCM-2017.