IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received April 23, 2020, accepted May 14, 2020, date of publication May 22, 2020, date of current version June 8, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2996770

Colorectal Disease Classification Using Efficiently Scaled Dilation in Convolutional Neural Network

SAHADEV POUDEL[1], YOON JAE KIM[2], DUC MY VO[1], AND SANG-WOONG LEE[1], (Senior Member, IEEE)

[1]Department of Software, Gachon University, Seongnam 13557, South Korea
[2]Department of Internal Medicine, Gachon University Gil Medical Center, Incheon 21565, South Korea

Corresponding author: Sang-Woong Lee (slee@gachon.ac.kr)

This work was supported by the Gyeonggi-do Regional Research Center (GRRC) program of Gyeonggi province [GRRC-Gachon2017(B01), Analysis of behavior based on senior life log] and the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (2020R1A2C1011708).

ABSTRACT Computer-aided diagnosis systems developed by computer vision researchers have helped doctors to recognize several endoscopic colorectal diseases more rapidly, which allows appropriate treatment and increases the patient's survival ratio. Herein, we present a robust architecture for endoscopic image classification using an efficient dilation in Convolutional Neural Network (CNNs). It has a high receptive field of view at the deep layers in increasing and decreasing dilation factor to preserve spatial details. We argue that dimensionality reduction in CNN can cause the loss of spatial information, resulting in miss of polyps and confusion in similar-looking images. Additionally, we use a regularization technique called DropBlock to reduce overfitting and deal with noise and artifacts. We compare and evaluate our method using various metrics: accuracy, recall, precision, and F1-score. Our experiments demonstrate that the proposed method provides the F1-score of 0.93 for Colorectal dataset and F1-score of 0.88 for KVASIR dataset. Experiments show higher accuracy of the proposed method over traditional methods when classifying endoscopic colon diseases.

INDEX TERMS Colorectal image classification, colon disease classification, colon disease classification with CNN.

I. INTRODUCTION

Colorectal cancer is one of the most common and deadly cancers worldwide. Colon diseases like adenoma, adenocarcinoma, Crohn's disease, ulcerative colitis, adenocarcinoma, and adenoma—are considered as significant factors in the evolution of cancer [1]. According to the American Cancer Society, nearly 80,000 people died in 2016 because of different types of cancers related to the gastrointestinal tract [2]. An accurate diagnosis of each disease is essential for early detection and effective treatment of colorectal cancer, which improves the patient's survival rate. Advances in technology make it possible for artificial intelligence (AI)-based computer vision approaches to assists doctors in colorectal classification tasks. Over the decades, research on artificial intelligence in medical imaging has been ongoing and shows its effectiveness in that particular domain [3], [4]. Other studies included an automatic classification of breast cancer [5]; skin cancer [6], detection of gastric cancer [7]; hookworms [8], and recognition of brain tumors [9], [10].

Traditional machine learning focused on handcrafted feature-based methods, which rely on image color, shape, and texture information. These methods followed the same approach which requires a feature extraction process and the use of several classifiers for the classification process. However, feature extraction was difficult because of a lack of illumination, blurring, variations in viewpoint, and even colon insufflation.

In recent years, inspired by the great success of deep learning (DL) in computer vision [11]–[13], the interest of applying deep learning to endoscopic image analysis is increasing. Despite this, obtaining a large amount of balanced data remains challenging in the medical field. Even though the transfer learning approach can be applied to solve the above problem [14]–[16], it still suffers from some serious issues. One example is the excessive use of downsampling approaches in the deep layers of a pre-trained network which may work well for natural image classification such as ImageNet dataset [17] but not in the medical domain due to its high intra-class variance and low inter-class variance in different classes. We argue that the small feature maps at the deeper layers of the network contain abstract

The associate editor coordinating the review of this manuscript and approving it for publication was Chun-Wei Tsai.

information which is not enough to represent endoscopic features, thus confound in small polyps and similar-looking images. We think that the bigger size of outputs will represent the features more explicitly, which would improve the classification. Another example is the possibility of occurrence of overfitting when feeding a fewer number of samples to pre-trained CNNs; it may learn the noise and detail in the data, which can negatively impact the performance on unseen data. On the one hand, gaining a high accuracy in similar-looking classes with the limited dataset is difficult but a key demand. On the other hand, the utilization of a pre-trained network without effective use of regularization methods are very prone to overfitting. This greatly limits the use of traditional CNNs in the medical world. A small marginal classification error in endoscopic images can lead to a bad experience in the medical domain. For example, the two diseases Crohn's and ulcerative colitis, share similar features that are characterized by chronic inflammation of the digestive tract, and a spurious prediction of such disease is not acceptable at any cost in clinical settings. It is therefore desired to conceive more effective endoscopic image classification architectures that can effectively recover the fine details in the medical images.

To address the need for more accurate classification in the medical domain, we propose a novel method that increases the receptive field of view in the deep layers using dilated convolution effectively. The root presumption behind our architecture is that the model can capture more fine details with the use of dilation at the last layers and helps in increasing performance when high-resolution feature maps are passed to the classification layer. The dilation is added in increasing and decreasing order to get rid of gridding artifact problems. We deduce that increasing use of dilation factors cannot aggregate important spatial features of small polyps and similar-looking images and therefore is detrimental to images of such classes. Similarly, benefited from the regularization method, a DropBlock [18] is added after each dilated convolution to deals with noise, artifacts and reduce overfitting. It drops out the adjacent region of a feature map together which forces the network to look elsewhere to fit the data and hence, helps in the regularization of a model. Since the dataset suffers from various artifacts like specular reflection, artificial devices, motion blur, we think that use of the DropBlock method can handle such artifacts.

Our paper begins with an introduction and motivation for the proposed approach. Next, we present related studies on the endoscopic classification of colorectal diseases (in Section II) with a summary of our contributions. Section III presents a comprehensive description of our proposed method for the classification of several colorectal diseases. In Section IV, we show a collection of data and performance metrics and network training. Section V display our experimental results and performance analysis of our proposed system. Section VI interprets and describes the significance of our findings. Finally, Section VII concludes the paper with a summary of our contributions.

II. RELATED WORK

In this section, we describe various feature extraction and classification methods, including handcrafted feature-based methods and deep learning-based algorithms that have been proposed to classify endoscopic images of colorectal diseases.

A. COLONIC POLYPS

Colonic polyps are considered as a major precursor of colon cancer. In an early study, Häfner *et al.* [19] introduced texture analysis methods based on local fractal dimension (LFD) for the classification of colonic polyps. They proposed three LDF-based approaches that additionally extracted shape and gradient information of the image to improve classification; these methods were tested on different datasets. Next, a filter bank-based texture analysis method was proposed for the classification of colonic polyps [20]. Different types of polyps were differentiated using the filter masks of the filter bank. M. Hafner *et al.* proposed a novel color texture operator that was based on a noise-robust local binary pattern variant for an automatic classification of endoscopic images [21]. They quantified the similarity of neighboring pixels by constructing a color vector field from an image and used k-nearest neighbors classifier for classification. Wimmer *et al.* [22] tested several wavelet-based approaches for 11 endoscopic polyp databases, proposed three wavelet-based feature extraction approaches, and found them acceptable for an automatic classification of colonic polyps. Tamaki *et al.* [23] proposed a local feature-based recognition system: a bag-of-visual words representation of local features followed by the support vector machine (SVM) classifier. In [24], they integrated a Gabor filter and monogenic local binary pattern to generate a new feature that represented shape and edge information at multiresolution while preserving color information. Consequently, linear discriminant analysis was used to reduce the feature dimensions, and SVM was used as a classifier. Stehle *et al.* [25] proposed a classification algorithm for colonic polyps: they implemented two segmentation algorithms, and the obtained features were used to classify the polyps.

Recently, CNNs have been used instead of handcrafted features for automatic feature extraction and classification [37]. Pogorelov *et al.* [15] combined deep neural networks, information retrieval, and analysis of global and local image features for multiclass classification, detection, and localization of various gastrointestinal diseases. Zhang *et al.* [16] proposed a transfer learning approach by using the features learned from non-medical datasets using deep CNN; subsequently, they used low-level features to detect and localize colorectal polyps. Shin and Balasingham [36] have shown that CNN outperformed handcrafted feature-based methods after comparing them on three public polyp databases. Nadeem *et al.* [32] integrated texture and deep learning features for the classification of gastrointestinal diseases. Urban *et al.* [33] designed a deep CNN to detect polyps and

TABLE 1. Comparison and weaknesses of previous approaches.

Publications	Method	Classification type	Dataset	Accuracy	Weaknesses
M. Hafner et al. [21]	Multiscale novel color texture + local binary pattern (LBP)	polyp	private	85.3%	-Requires handcrafted features
D. Mahapatra et al. [26]	Supervised learning + intensity and texture	crohn's	private	88.9%	-Requires handcrafted features
Z. Wei et al. [27]	Gabor filter banks + K-means clustering + histogram	colitis	private	-	-Requires handcrafted features
M. Hafner et al. [19]	Local fractal dimension	polyp	private	88.2%	-Requires handcrafted features
S. S. Ahmed et al. [28]	Neuro-fuzzy model	crohn's	private	97.67%	
E. Mossotto et al. [29]	Combined unsupervised ML	crohn's/colitis	private	83.3%	-Reliance on endoscopic and histological data
A. Alammari et al. [30]	Convolutional neural network	colitis	private	-	-Failure in images with similar features
G. Wimmer, A. Vecsei, M. Hafner, A. Uhl [31]	CNN + SVM	polyp/colitis	KSAVIR (public)	92.5%	-Network complexities -
M. Zaid et al. [32]	Ensemble of texture and CNN features	polyp/colitis	private	83% -	-Requires handcrafted features -Use of ensembling techniques
G. Urban et al. [33]	Deep CNN	polyp		96%	-Experiment on less number of class
R. W. Stidham et al. [34]	Deep CNN	colitis	private	-	-Lack of diverse dataset -Less reliable, overfitting
T. Ozawa et al. [35]	CAD + CNN	colitis	private	-	-Use of imbalanced dataset
R. Zhang et al. [16]	CNN + SVM	polyp	public	85.9%	-Less number of images -Lack of diverse datasets
Y. Shin and L.Balansingham [36]	3-layer CNN	polyp	public	91.26%	-less number of classes and images

evaluated the results with an expert colonoscopist. All polyps identified in the expert review were also detected by their proposed method.

Wimmer *et al.* [31] applied three pre-trained CNN architectures to endoscopic image databases, and SVM was subsequently used to classify colonic polyps and celiac disease. They concatenated and combined the features from several layers and experimented with classification. Their approach outperformed other CNN-based approaches.

B. CROHN's DISEASE AND ULCERATIVE COLITIS

Mahapatra *et al.* [38] proposed a supervised learning approach for automatic identification and localization of the regions affected by Crohn's disease in abdominal magnetic resonance volumes. They used intensity statistics, texture anisotropy, and shape asymmetry of the 3D regions as features to distinguish between normal and affected regions. In [26], D. Mahapatra *et al.* performed similar tasks with the use of low-level features such as intensity and texture. Wei *et al.* [27] used a visual codebook to accurately detect colitis on contrast-enhanced computed tomography scans. Ahmed *et al.* [28] defined a neuro-fuzzy-based approach that combined a backpropagation neural network-fuzzy classifier with a neuro-fuzzy model to diagnose Crohn's disease. They used factor analysis as a dimensionality reduction technique and performed experiments on different levels of the fuzzy partition.

Mossotto *et al.* [29] proposed three unsupervised ML models that used endoscopic data only, histological data only, and combined endoscopic/histological data, achieving an accuracy of 71, 76.9, and 82.7 percent, respectively. Han *et al.* [39] developed a novel pathway-based approach that used genes to calculate individualized pathway scores for the classification of ulcerative colitis and Crohn's disease. Pogorelov *et al.* [40] presented a dataset named "KVASIR," where different diseases were classified using global features, deep CNNs, and deep transfer learning. Alammari *et al.* [30] proposed an approach that used endoscopic domain knowledge and a deep CNN to classify the severity of ulcerative colitis. Stidham *et al.* [34] found that the accuracy of a deep CNN was comparable to experienced human reviewers for the classification of endoscopic severity of ulcerative colitis. Ozawa *et al.* [35] showed the robustness of a GoogLeNet CNN architecture based on a computer-aided diagnosis (CAD) system for identifying endoscopic inflammation severity in ulcerative colitis. Maeda *et al.* [41] developed a CAD system for predicting persistent histologic inflammation associated with ulcerative colitis.

C. LIMITATIONS OF RELATED WORK

Table 1 summarizes the problems in existing classification approaches. Previous methods have at least one of the following weaknesses:

- Dependence on a fixed set of handcrafted features which requires a deep knowledge about the image characteristics [19], [21]–[27]. They relies on texture analysis where a limited set of local descriptors computed from an image is fed into a classifiers like SVM,Random Forests etc. Despite a good level of accuracy in some works, these techniques have limitations on generalization and transfer capabilities in inter-dataset variability.
- Experimented on less number of classes [16], [33], [36]. The work on [34] were tested on less diverse dataset.
- Reliance on endoscopic and histological data which limits the practical utility of these algorithms [29], [41] since histological data might not be available in all scenarios.

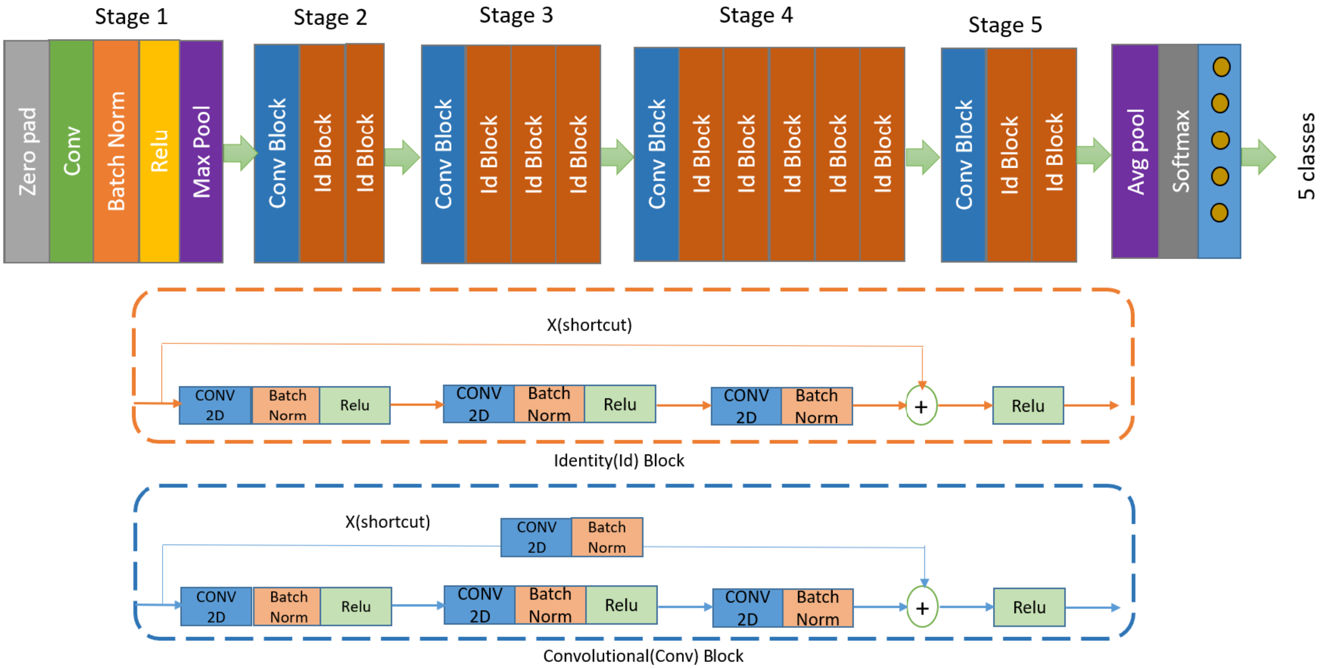

FIGURE 1. Overview of original ResNet50 architecture [42]. At stage 1, the feature map size is downsampled by a convolutional layer with strides = 2, which is followed by Batch normalization and Relu layer. Within each stage, the number of filters used by the layers is the same. Each stage has convolutional(Conv) block and Identity(Id) block. The identity block contains three sets of a convolutional layer, and the Convolutional block has one extra layer to match the input and output dimension.

- Incognizant about the features the network learned during the training process [15], [30], [31], [33]–[36], [40].

Contrary to previous works, our approach does not rely on handcrafted features and histological data but uses a deep learning-based neural network using efficient dilation with an effective regularization method approach for endoscopic image classification. Further, The use of a diverse and large number of classes and images on the proposed method makes the model more reliable. In this paper, we present the following major contributions:

1) We increase the receptive field of view in the deep layers of the network with the efficient use of dilated convolution to preserve the spatial information. We utilize the dilation factor in increasing and decreasing order to aggregate the spatial details of tiny features like polyps.
2) We further validate the use of a regularization technique called DropBlock to avoid overfitting and handles noises and several artifacts like specular reflection, artificial devices, motion blur.
3) Finally, we evaluate our proposed deep neural network on our colorectal dataset that includes five classes, and we additionally evaluate it on another endoscopic KVASIR dataset [40]. We show that our approach is promising for endoscopic image classification.

III. METHODOLOGY

Our CNN model for selecting the features from endoscopic images is based on the transfer learning approach: rather than training a new CNN architecture, we reuse a pre-trained network. It is widely known that the features extracted from the activation of a CNN trained in a fully supervised manner for a large-scale object recognition task can be repurposed for a novel generic task. Moreover, our training set contains only a few hundreds of images which is insufficient for state-of-the-art CNN architectures that require millions of parameters to train. Tajbakhsh *et al.* [14] demonstrated that the use of a pre-trained CNN with adequate fine-tuning outperformed or, in the worst case, performed similarly as a CNN trained from scratch. Therefore, we employed the last layer fine-tuning on endoscopic data set and initialized the ImageNet pre-trained weights for each model, and the last fully connected layer is updated continuously. We use the ResNet50 architecture as a baseline model. From extensive experiments, we found that ResNet50 achieved better performance than

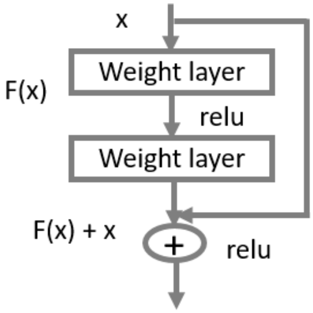

FIGURE 2. Residual learning: a building block [42].

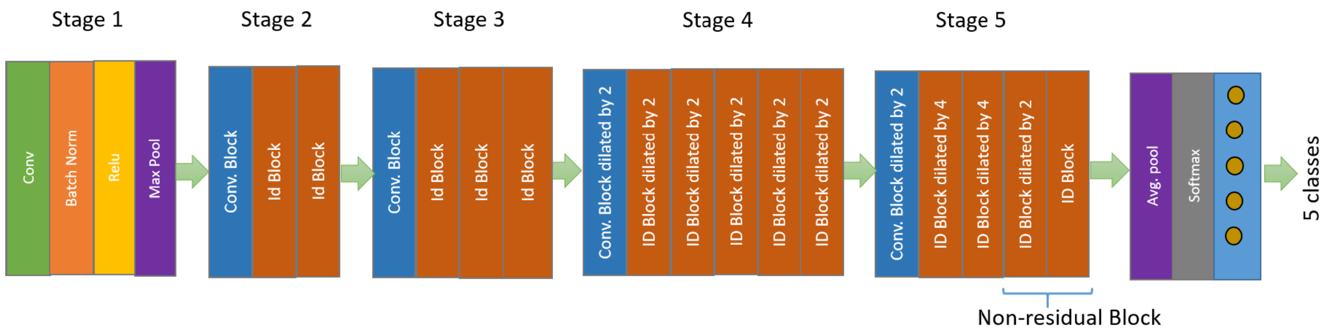

FIGURE 3. A proposed architecture. Layers at stage 4 and stage 5 are dilated, and two non-residual blocks are added to the end. Striding in the first block of stage 4 and stage 5 is removed, and remaining blocks are dilated, explained in section IV. Every convolutional layer is followed by the DropBlock regularization method at these stages.

Layer Name	Conv1	Conv2_x		Conv3_x	Conv4_x	Conv5_x					Fc
Original ResNet	7*7,64, stride 2 output size 112 * 112	3*3 max pooling, stride 2	(1*1,64 3*3,64 1*1,256) X 3 output size 56*56	(1*1,128 3*3,128 1*1,512) X 4 output size 28*28	(1*1,256 3*3,256 1*1,1024) X 6 output size 14*14	(1*1,512 3*3,512 1*1,2048) X 3				output size 7*7	Average pool, number of class fc, softmax
Proposed Method			(1*1,64 3*3,64 1*1,256) X 3 output size 56*56	(1*1,128 3*3,128 1*1,512) X 4 output size 28*28	(1*1,256 3*3,256 dilate 2 1*1,1024) X 6 output size 28*28	(1*1,512 3*3,512 dilate 2 1*1,2048) X 1	(1*1,512 3*3,512 dilate 4 1*1,2048) X 2	(1*1,512 3*3,512 dilate 2 1*1,2048) X 1	(1*1,512 3*3,512 dilate 1 1*1,2048) X 1	output size 28*28	

FIGURE 4. A detailed structure of the proposed method and the baseline model (ResNet50).

other existing CNN architectures [12], [43]–[46] in colorectal dataset.

Fig. 2 shows a basic residual network block with the inclusion of the identity connection. The residual block will learn the following function.

$$H(x) = F(x) + x, \tag{1}$$

where, $F(x)$ is represented by the stacked non-linear layers and x is an identity function. The ResNet50 architecture consists of five stages of blocks where small chunks of networks connected through skip or shortcut connections to form an extensive network. Two main types of blocks are used, depending mainly on whether the input/output dimensions are the same or different. When the input activation has the same size as the output activation, it is formulated as:

$$y = F(x, W_i) + x, \tag{2}$$

where, x and y indicates the input and output vectors of the layers considered. The function $F(x, W_i) + x$ represents the residual mapping to be learned. Fig. 1 shows an example of an identity block, where the upper path is the shortcut path and the lower path is the main path. Similarly, when the input and output dimensions do not match, we add a convolutional layer to the shortcut path by using the following formula

$$y = F(x, W_i) + W_s x \tag{3}$$

Usually, each identity block contains three sets of convolutional layers followed by batch normalization and the ReLU activation function. Similarly, the convolutional block includes the same number of layers with one extra convolutional layer added.

From our assumption, We need to keep the network from down-sampling approach and preserve complex spatial information to the last layers. We achieve this by providing dilation and removing down-sampling. We adopt a network architecture [47] which was designed for ImageNet Classification and make several modifications to fit the network for our purposes.

A. APPROACH

In this section, we describe how our proposed model learns and represents the endoscopic features of colon diseases effectively. For this, we use dilated convolutions with different atrous rate at the end of the layers. The proposed architecture consists of five groups of layers with convolution and identity blocks described in Section III. Let G^i be a group of layers, where $i = 1, \ldots ..5$. We denote the j^{th} layer in group i as G^i_j. Let f^i_j represent the filter associated with layer G^i_j. The output of G^i_j in the original model is

$$(G^i_j * f^i_j)(p) = \sum_{a+b=p} G^i_j(a) f^i_j(b). \tag{4}$$

We use dilated convolutions in the final two groups of convolutional layers. For Stage 4, we replace the convolution operators by dilated convolutions with an atrous rate of 2 for all layers of the block.

$$(G^4_j * 2f^4_j)(p) = \sum_{a+2b=p} G^4_j(a) f^4_j(b) \tag{5}$$

for all $j > 0$. In the first layer of the block in Stage 5, which is G^5_1, we perform the same transformation.

$$(G^5_j * 2f^5_j)(p) = \sum_{a+2b=p} G^5_j(a) f^5_j(b) \tag{6}$$

By analogy, we use a dilated factor of 4 in the remaining blocks of G^5_j:

$$(G^5_j * 4f^5_j)(p) = \sum_{a+4b=p} G^5_j(a) f^5_j(b) \tag{7}$$

for all $1 < j < 4$. Similarly, for a dilation factor of 2 in the fourth block of stage 5 layer,

$$(G^5_j * 2f^5_j)(p) = \sum_{a+2b=p} G^5_j(a) f^5_j(b) \tag{8}$$

for $3 < j < 5$. Then, a non-residual block having normal convolution is added at the end which is followed by the global average pooling layer (as in the original architecture), which decreases the output feature maps to a vector, and 1×1 convolution maps this vector to a vector that contains the prediction scores for all classes. The overall proposed and modified architecture is illustrated in Fig. 3 and pseudocode is shown in Algorithm 1.

The layer-wise details of the architecture are explained in Fig. 4, which exhibits information about each layer of both original and proposed architecture in sequential order. Our proposed model consists of fifty-seven layers: fifty-six convolutional layers followed by batch normalization that normalizes the feature map and an activation function called the rectified linear unit (ReLU). Because only the convolutions in the later layers of the networks are dilated, the shape and structure of earlier layers are the same. For both models, the first convolutional layer generates a feature map of size $112 \times 112 \times 64$ after applying 64 different filters of size $7 \times 7 \times 3$ over the input image of size 224×224. Then, a max-pooling layer is used, which processes the input feature map by applying a filter of size 3×3 pixels to generate the feature map of $56 \times 56 \times 64$. For the original model, downsampling is accomplished by the first 1×1 convolution layer with a stride of 2 in the layers of Stage 3, Stage 4, and Stage 5. However, in our proposed model, we set stride to 1 and replace the 3×3 convolution with 3×3 dilated convolution after Stage 3. we gradually increase and decrease the dilation rate in the convolutional layers of Stage 4 and Stage 5 and remove the residual connection at the final two layers. Then, the optimal feature vector of size $1 \times 1 \times 2048$ is generated after applying the global average pooling layer.

Algorithm 1 Overall Proposed Method

```
Initialize the ResNet50 Networks
ResNet ← load()
features,labels = get_batch(dataset)
model ← create_model()
**if** $mode == training$ **then**
    fitted-model ← model.fit(features,labels)
    **for** $e = 1 : epochs$ **do**
        **for** i = 1: $G^i$ **do** ▷ $G^i$ is different stages of layers.
            **if** $G^i_j == G^4_j$ **then** ▷ $G^i_j$ is $j^{th}$ layer in group $i$
                Execute equation number (5),
                Add DropBlock
            **end if**
            **if** $G^i_j == G^5_j$ **then**
                Execute equation number (6,7,8),
                Add DropBlock
            **end if**
        **end for**
    **end for**
model ← fitted-model()
**end if**
```

Algorithm 2 DropBlock Regularization Method

```
**Input:**output activations of a layer (A), $block\_size, \gamma$
Sample mask matrix $M$ randomly, where $M_{i,j} \sim Bernoulli(\gamma)$;
For each zero position $M_{i,j}$, expand a spatial square mask
    with the center being $M : M_{i,j}$, the width, height being
    $block\_size$ and set all the values of $M$ in the square to be
    zero.
Apply the mask: $A = A * M$
Normalize the features:$A = A * count(M)/count\_ones(M)$
```

The two non-residual blocks with decreasing dilation are added in the proposed method to overcome the problem of gridding artifacts. Gridding artifacts occur when a feature map has a higher-frequency content than the dilated convolution sampling rate. By doing this, the model does not allow propagation of gridding artifacts from the earlier layers. The converted network will generate the output of 28×28 after G^5 layers, which helps global average pooling layers to take more values. It helps the classifier to recognize the features that cover a tiny part in the given image.

Moreover, adding more non-residual blocks at the end layer increases the network size, which can cause overfitting or get stuck in poor local minima. Our limited dataset

TABLE 2. Details of colorectal dataset.

Actual Class	Training set		Validation set	Testing set
	Before Augmentation	After Augmentation		
Adenocarcinoma	474	7380	80	80
Adenoma	615	7380	80	80
Crohn's	403	7380	80	80
Ulcerative colitis	613	7380	80	80
Normal	610	7380	80	80

might not be helpful in this case. Also, the presence of high background noises and artifacts in the endoscopic images is one of the challenges encountered during classification. To overcome this problem, we utilize the effective use of DropBlock [18] method, which is beneficial to regularize convolutional networks. It drops an adjacent region of a feature map together, unlike Dropout [48], which drops out features randomly. We applied DropBlock in all blocks of stage 4 and 5 after each convolutional layers. The pseudocode is explained in algorithm 2. It has two main parameters which are block_size u and γ. The block_size u is the length of contiguous region to be dropped, while γ controls how many units to drop. We use the fixed size of u of *7*7* after the convolutional layer. Similarly, we compute γ by following formula:

$$\gamma = \frac{(1 - keep\_prob)(v^2)}{u^2(v - u + 1)^2}, \tag{9}$$

where *keep_prob* is the probability of keeping an every activation unit in dropout. We sample the initial binary mask with the Bernoulli distribution with a mean of $1 - keep\_prob$. v is the size of a feature map, and $(v - u + 1)^2$ indicates the size of the valid seed region. In our experiments, we use $keep\_prob = 0.9$ in all layers and compute the value of γ.

IV. EXPERIMENTAL PROTOCOL

A. DATA COLLECTION

1) COLORECTAL DATA

The dataset was provided by Gill Hospital, South Korea, and it contains five classes of 3,515 endoscopic colorectal disease images: 634 with adenocarcinoma, 775 with adenoma, 563 with Crohn's disease, 773 with ulcerative colitis, and 770 normal images. The original image sizes range from 400×400 to 2000×2000 pixels. Therefore, the images are resized according to the requirements of the CNNs architecture. The image data is normalized with the default properties required for each architecture. We perform data augmentation to increase the number of images before training the networks due to the small amount of available data. Originally, the class is imbalanced, and augmentation is done in such a way that the minority class is augmented more to make a balanced dataset. This is a standard solution to reduce overfitting during the training. Several augmentation techniques such as flipping, scaling, rotating, zooming, contrast normalizing, and shearing was used. Each image was first rotated at a different angle, and each rotated image was flipped each time (horizontally and vertically) and zoomed. Before augmentation, we split the total dataset and separated 80 images for validation and testing purposes, and the remaining images belonged to training. The details of a colorectal dataset are presented in Table 2.

2) KVASIR DATASET

KVASIR Dataset includes 4000 endoscopic gastrointestinal diseases and comprises eight different classes, each containing 500 images. The dataset consists of several sets of images in each category, including anatomical landmarks (such as Z-line, pylorus, or cecum) and pathological findings (such as esophagitis, polyps, or ulcerative colitis). Some sets are related to the removal of lesions, including dyed and lifted polyps and dyed resection margins. Images with different resolutions from 720 × 576 to 1920 × 1072 pixels are included in the dataset. Pogorelov *et al.* [40] performed a baseline evaluation of these datasets with three different approaches: classification using global features, deep convolutional neural networks, and deep transfer learning. We split the dataset into a 50:50 ratio to make a fair comparison with the original paper. We will compare the results of our proposed method with these existing approaches.

B. PERFORMANCE METRICS

We use performance evaluation metrics such as accuracy (ACC), recall, precision, and F1-score to evaluate classifiers, computed as follows:

$$ACC = \frac{(P_T + N_T)}{(P_T + N_T + P_F + N_F)}, \tag{10}$$

$$recall = \frac{P_T}{P_{T+N_F}}, \tag{11}$$

$$precision = \frac{N_T}{N_{T+P_F}}, \tag{12}$$

$$F1 - score = 2 * \frac{recall * precision}{recall + precision}, \tag{13}$$

where ACC and $F1$ are accuracy and F1-score, respectively; P_T and N_T are the number of true positives and true negatives, respectively; P_F and N_F are the number of false positives and false negatives, respectively. Specifically, ACC is the proportion of correctly classified samples. Precision is the proportion of true negatives that are correctly classified. A recall is the proportion of true positives that are correctly classified. The F1-score is the harmonic average of precision and recall.

TABLE 3. F-score (F1), precision, recall for the evaluated different CNNs architecture on the colorectal dataset.

Model	Class	precision	recall	F1
VGG16 [12]	adenocarcinoma	0.88	0.91	0.90
	adenoma	0.94	0.91	0.92
	crohn's disease	0.85	0.69	0.76
	normal	0.95	0.94	0.94
	ulcerative colitis	0.79	0.94	0.86
		0.88	0.87	0.87
InceptionResnetv2 [43]	adenocarcinoma	0.83	0.96	0.89
	adenoma	0.93	0.86	0.90
	crohn's disease	0.93	0.78	0.84
	normal	0.99	0.90	0.95
	ulcerative colitis	0.82	0.96	0.89
		0.90	0.89	0.89
Xception [44]	adenocarcinoma	0.89	0.89	0.89
	adenoma	0.90	0.86	0.88
	crohn's disease	0.88	0.72	0.79
	normal	0.97	0.93	0.95
	ulcerative colitis	0.74	0.94	0.83
		0.87	0.86	0.86
ResNet [42]	adenocarcinoma	0.90	0.94	0.92
	adenoma	0.94	0.91	0.92
	crohn's disease	0.91	0.79	0.85
	normal	0.96	0.97	0.97
	Ulcerative colitis	0.83	0.93	0.88
		0.91	**0.91**	**0.91**
DenseNet [45]	adenocarcinoma	0.90	0.91	0.91
	adenoma	0.94	0.93	0.93
	crohn's disease	0.90	0.76	0.82
	normal	0.97	0.95	0.96
	ulcerative colitis	0.81	0.95	0.87
		0.90	0.90	0.90
NasNet [46]	adenocarcinoma	0.84	0.96	0.90
	adenoma	0.93	0.88	0.90
	crohn's disease	0.93	0.79	0.85
	normal	0.99	0.89	0.94
	ulcerative colitis	0.82	0.96	0.89
		0.90	0.89	0.90

TABLE 4. F-score(F1), precision, recall for the evaluated different existing approaches on colorectal dataset.

Model	Class	precision	recall	F1
R. Zhang et al. [16]	adenocarcinoma	0.78	0.81	0.80
	adenoma	0.82	0.94	0.88
	crohn's disease	0.79	0.79	0.79
	normal	0.99	0.98	**0.99**
	ulcerative colitis	0.79	0.84	0.81
		0.83	0.87	0.85
Y. Shin and I. Balansingham [36]	adenocarcinoma	0.79	0.90	0.84
	adenoma	0.89	0.95	0.92
	crohn's disease	0.93	0.54	0.73
	normal	0.93	0.95	0.94
	ulcerative colitis	0.64	0.90	0.75
		0.83	0.84	0.83
W. Ryan et al.[34]	adenocarcinoma	0.89	0.91	0.90
	adenoma	0.94	0.91	0.92
	crohn's disease	0.86	0.75	0.80
	normal	0.97	0.94	0.96
	ulcerative colitis	0.81	0.94	0.87
		0.894	0.89	0.89
Proposed Method	adenocarcinoma	0.96	0.94	**0.95**
	adenoma	0.95	0.97	**0.96**
	crohn's disease	0.92	0.82	**0.87**
	normal	0.99	0.97	0.98
	ulcerative colitis	0.84	0.94	**0.89**
		0.932	0.928	**0.93**

TABLE 5. Accuracy (Acc), f-score (F1), recall, precision for the kvasir dataset.

Method	Acc	Precision	Recall	F1
6 Layer CNN	0.914	0.661	0.640	0.651
3 Layer CNN	0.959	0.589	0.408	0.453
Inception v3 TFL	0.924	0.698	0.689	0.693
2 GF Random Forest	0.928	0.713	0.715	0.711
2 GF Logistic Model	0.926	0.706	0.707	0.705
6 GF Random Forest	0.933	0.732	0.732	0.727
6 GF Logistic Model Tree	0.937	0.748	0.748	0.747
Proposed Method	**0.957**	**0.868**	**0.922**	**0.88**

C. NETWORK TRAINING

The implementation is based on Keras, and the backend is TensorFlow.The training set is used to train the model and learn the parameters. The validation set is used to optimize the model and test it during the training: to automatically adjust the learning rate and decide whether to stop early according to the test performance of a given training step. The test set is used to evaluate the recognition and generalization ability of the proposed model.

We initialize the pre-trained weights of ResNet50 and use stochastic gradient descent with a batch size of 16. The learning rate starts from 0.001 and is divided by 10 when the patience level exceeds 8. We use a weight decay of 0.0001 and a momentum of 0.9 without an accelerated gradient.

V. EXPERIMENTAL RESULTS AND EVALUATION

A. PERFORMANCE EVALUATION OF COLORECTAL DATASET

In this section, we will compare the performance of the proposed method with the existing related methods which were used in endoscopic image classification. Due to limited works on colorectal diseases using deep learning, we compare the proposed work with the methods used for other similar tasks.

Table 3 presents the result when different fine-tuned CNNs were trained on the colorectal dataset for classification. The experiments are performed with the same parameters and the same number of augmented and validation sets. All the architectures achieved similar results, but the best results are obtained by ResNet50. From Table 4, we can observe the performance of each method evaluated by the F1-score.

Our model significantly outperforms the existing methods by a vast margin achieving 0.93 F1-score. The Zhang *et al.* [16] achieves good accuracy in normal class which shows its discriminative capability between normal and disease class but failed to achieve a similar result on similar-looking disease classes. Shin and Balansingham [36] with three layers achieves 0.836 F1-score indicating that it is not deep enough to learn the complex patterns of the images resulting in poor performance. Meanwhile, Stidham *et al.* [34] with 159 layers achieves 0.89 F1-score that shows its powerful discriminative capability but achieves similar result with the other existing CNNs method.

B. PERFORMANCE EVALUATION OF KVASIR DATASET

The results obtained are presented in Table 5 with the metrics like *precision*, *recall*, $F1$, *Accuracy* measures estimated from the TP, FP, and TN and FN cases. We compare our proposed

TABLE 6. Ablation studies for the dilated rate at stage 4 and stage 5. We evaluated the proposed method on the given values.

Method	Dilation rate at stage 4	Dilation rate at stage 5	DropBlock	precision	recall	F1
Baseline model (ResNet50)	1	1	-	0.91	0.91	0.91
ResNet50	1	2	-	0.91	0.92	0.91
	2	2,4	-	0.88	0.91	0.89
	2	2,4	✓	0.88	0.92	0.90
	2	2,4,2	-	0.91	0.90	0.90
	2	2,4,2	✓	0.90	0.93	0.91
	2	2,4,2,1	-	0.93	0.92	0.92
Proposed Method	2	2,4,2,1	✓	0.932	0.928	0.93

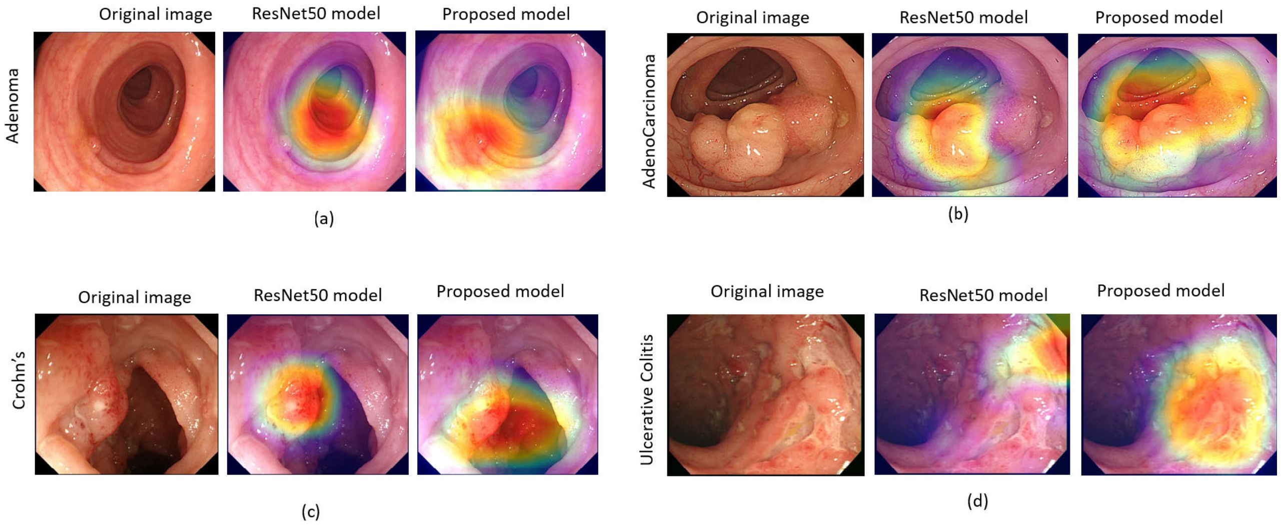

FIGURE 5. Comparison of CAMs generated from the proposed and the baseline ResNet50 method. The proposed method highlighted the specific regions which were misclassified by the baseline model.

method with the baseline performance model stated in the research. Their baseline model includes: classification using global features (GF), deep learning convolutional neural networks (CNN), and transfer learning in deep learning (TFL). It can be noticed that the proposed methodology achieved 95.7 % accuracy on these datasets with F-score value of 0.88, which is slightly better than the 2 Layer CNN and 3% better than the Inception model using TFL.

C. ABLATION STUDY

We investigate the effectiveness of our contributions by comparing our full model with the baselines based on the same experimental setting. For each ablation experiment, we omit different dilation rates at the convolutional layers of stage 4 and 5 block. The results obtained are displayed in Table 6. We report the F1, Precision, and Recall score for each experimented values.

From the results reported in Table 6. we can draw the following conclusions: 1). Simply adding dilated convolution at the end layers does not improve the classification performance; instead, it will only worsen the performance. Similarly, the addition of a convolutional block with the dilation rate of 2 at stage 4 does not make any changes. Further, the F1-score drops from 0.91 to 0.89 and 0.90 when significant changes were made in the dilation rate (4th and 6th row) at stage 4 and stage 5, showing that the network suffers from the gridding artifact problem. 2). The F1-score of the network at the complementary branch (5th and 7th row) is higher than the corresponding network in the 4th and 6th row, demonstrating the effectiveness of using DropBlock as a regularization method. 3). The network with increasing and decreasing dilation rates at the end layers improve the F1-score of the baseline network by 0.91 to 0.92. Further, utilizing of DropBlock regularization method enhances the network performance by 0.92 to 0.93. It shows that our approach has a better discriminative capability of identifying a small polyp and recognizing similar-looking images.

VI. DISCUSSION

Table 4 shows that our proposed novel deep model has achieved the best classification performance on the provided dataset. This study's results showed that removing downsampling and preserving features at the last blocks of the CNN increases performance with a 92.8% recall rate and a 93.2% precision rate.

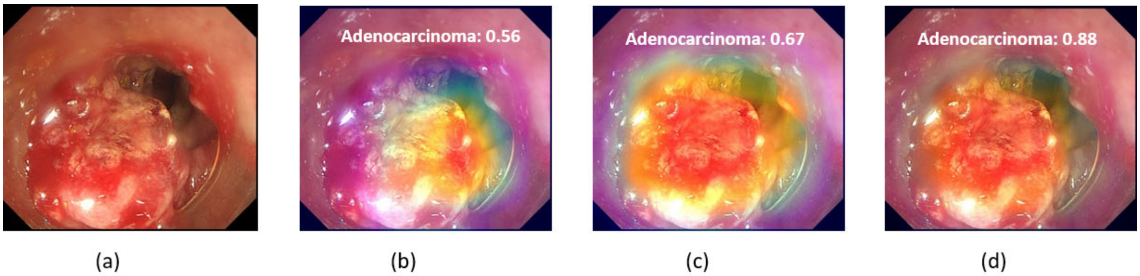

FIGURE 6. Class activation mapping (CAM): (a) original image (b) original ResNet50 model, (c) Dilated ResNet50 trained without DropBlock, (d) Dilated ResNet50 trained with DropBlock (Proposed Method). A proposed method tends to focus on the affected part only and less sensitive to noise, unlike (c).

The result from Table 3 indicated that the applying of transfer learning on medical datasets is not always beneficial in the medical domain. As it uses a progressive down-sampling approach in the CNN, it is not helpful for those datasets with high interclass similarity and intraclass variation.

Fig. 5 demonstrates the type of endoscopic images of different classes where the best-performing model ResNet50 is failed. It is observable that when the size of the polyps is very small in the endoscopic colon images, the network could not classify it because of the downsampling approach used in the existing CNNs. Fig. 5a shows that *adenoma* is misclassified as *normal* because of loss of spatial information due to the continuous reduction in resolution of images, which represents the tiny feature maps of size 7 × 7 in the end. Similarly, as the learned features are more class-specific at deep layers, the common occurrence of similar features of images between different classes might make the classification process more difficult. For example, some *adenoma's* which are a polyp tends to progress towards *adenocarcinoma*, and they might share a similar shape with continuous inflammation. Fig. 5b shows that the model confuses *adenocarcinoma* with polyps. In the last two sets of images in Fig. 5, the model detects only some patterns of each class, and the network uses such features during the classification process. But our method preserves information until to end layers. It confirms that deep models with effective use of dilated convolutional layers at the end have an advantage in classifying the endoscopic images, over the fine-tuning of a state-of-the-art CNN architecture and several other methodologies.

One benefit of the proposed method is to tackle with the noises and artifacts present in the image. Fig. 6 shows the significance of using the DropBlock regularization method with the dilated convolution. The probability score is increased from 56% to 88%, indicating that the proposed method focuses and covers more specific and essential regions and is less sensitive towards noise and artifacts.

Our method showed a powerful ability to extract useful features from the endoscopic images. We observed the features learned by the CNNs at the last layers using the class activation map approach [49] when the images are hard to distinguish, the other methods produced a large fluctuation in accuracy rates. With the proposed method, it achieved better results for similarly looking images. Additionally, Table 5 shows that the proposed method achieved a high F1-score of 0.88 with 92% recall rates in the KVASIR dataset, which indicates the high capacity of recognizing disease class. Our proposed convolutional neural network is more accurate and stable than other popular traditional and deep models for endoscopic image classification.

VII. CONCLUSION

In this paper, we investigated the use of deep learning for colorectal endoscopic image classification. We showed that the features represented by the layers before the global average pooling are insufficient because of the use of excessive downsampling, which causes loss of spatial information. We applied an efficient technique to preserve the spatial information at the end of layers: specifically, using dilated convolutional layers in increasing and decreasing order. Besides, further use of the DropBlock regularization method at the deeper stages attained specific regions with less sensitivity towards noise and artifacts. We observed an improvement in classification, which proved that the proposed model captured more detailed and tiny differences between similar-looking images. Finally, with extensive experiments and comparisons on the KVASIR dataset, we demonstrated that our proposed deep convolutional neural network has a superior performance in endoscopic image classification. In our future work, we will further employ our novel neural network architecture to handle other endoscopic image classification problems. We also plan to extend our innovative approach by using earlier feature layers and deep features with dilated convolution to tackle image classification problems in other domains.

REFERENCES

[1] S. Hassanpour, B. Korbar, A. Olofson, A. Miraflor, C. Nicka, M. Suriawinata, L. Torresani, and A. Suriawinata, "Deep learning for classification of colorectal polyps on whole-slide images," *J. Pathol. Informat.*, vol. 8, no. 1, p. 30, 2017.

[2] D. Braithwaite *et al.*, *American Cancer Society: Cancer Facts and Figures 2016*. Atlanta, GA, USA: American Cancer Society, 2016.

[3] M. Owais, M. Arsalan, J. Choi, and K. R. Park, "Effective diagnosis and treatment through content-based medical image retrieval (CBMIR) by using artificial intelligence," *J. Clin. Med.*, vol. 8, no. 4, p. 462, Apr. 2019.

[4] F. Amato, A. López, E. M. Peña-Méndez, P. Vaňhara, A. Hampl, and J. Havel, "Artificial neural networks in medical diagnosis," *J. Appl. Biomed.*, vol. 11, no. 2, pp. 47–58, 2013.

[5] D. M. Vo, N.-Q. Nguyen, and S.-W. Lee, "Classification of breast cancer histology images using incremental boosting convolution networks," *Inf. Sci.*, vol. 482, pp. 123–138, May 2019.

[6] H. Takiyama, T. Ozawa, S. Ishihara, M. Fujishiro, S. Shichijo, S. Nomura, M. Miura, and T. Tada, "Automatic anatomical classification of esophagogastroduodenoscopy images using deep convolutional neural networks," *Sci. Rep.*, vol. 8, no. 1, p. 7497, Dec. 2018.

[7] 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, Jul. 2018.

[8] 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.

[9] B. Li and M. Q.-H. Meng, "Tumor recognition in wireless capsule endoscopy images using textural features and SVM-based feature selection," *IEEE Trans. Inf. Technol. Biomed.*, vol. 16, no. 3, pp. 323–329, May 2012.

[10] S. Sawant and M. Deshpande, "Tumor recognition in wireless capsule endoscopy images," *Int. J. Comput. Sci. Netw. Secur.*, vol. 15, no. 4, p. 85, 2015.

[11] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet large scale visual recognition challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, Dec. 2015.

[12] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: http://arxiv.org/abs/1409.1556

[13] K. Yun, J. Park, and J. Cho, "Robust human pose estimation for rotation via self-supervised learning," *IEEE Access*, vol. 8, pp. 32502–32517, 2020.

[14] N. Tajbakhsh, J. Y. Shin, S. R. Gurudu, R. T. Hurst, C. B. Kendall, M. B. Gotway, and J. Liang, "Convolutional neural networks for medical image analysis: Full training or fine tuning?" *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1299–1312, May 2016.

[15] K. Pogorelov, M. Riegler, S. L. Eskeland, T. de Lange, D. Johansen, C. Griwodz, P. T. Schmidt, and P. Halvorsen, "Efficient disease detection in gastrointestinal videos—Global features versus neural networks," *Multimedia Tools Appl.*, vol. 76, no. 21, pp. 22493–22525, Nov. 2017.

[16] 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 Informat.*, vol. 21, no. 1, pp. 41–47, Jan. 2017.

[17] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255.

[18] G. Ghiasi, T.-Y. Lin, and Q. V. Le, "DropBlock: A regularization method for convolutional networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2018, pp. 10727–10737.

[19] M. Häfner, T. Tamaki, S. Tanaka, A. Uhl, G. Wimmer, and S. Yoshida, "Local fractal dimension based approaches for colonic polyp classification," *Med. Image Anal.*, vol. 26, no. 1, pp. 92–107, Dec. 2015.

[20] G. Wimmer, A. Uhl, and M. Hafner, "A novel filterbank especially designed for the classification of colonic polyps," in *Proc. 23rd Int. Conf. Pattern Recognit. (ICPR)*, Dec. 2016, pp. 2150–2155.

[21] M. Häfner, M. Liedlgruber, A. Uhl, A. Vécsei, and F. Wrba, "Color treatment in endoscopic image classification using multi-scale local color vector patterns," *Med. Image Anal.*, vol. 16, no. 1, pp. 75–86, Jan. 2012.

[22] G. Wimmer, T. Tamaki, J. J. W. Tischendorf, M. Häfner, S. Yoshida, S. Tanaka, and A. Uhl, "Directional wavelet based features for colonic polyp classification," *Med. Image Anal.*, vol. 31, pp. 16–36, Jul. 2016.

[23] T. Tamaki, J. Yoshimuta, M. Kawakami, B. Raytchev, K. Kaneda, S. Yoshida, Y. Takemura, K. Onji, R. Miyaki, and S. Tanaka, "Computer-aided colorectal tumor classification in NBI endoscopy using local features," *Med. Image Anal.*, vol. 17, no. 1, pp. 78–100, Jan. 2013.

[24] Y. Yuan and M. Q.-H. Meng, "A novel feature for polyp detection in wireless capsule endoscopy images," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Sep. 2014, pp. 5010–5015.

[25] T. Stehle, R. Auer, S. Gross, A. Behrens, J. Wulff, T. Aach, R. Winograd, C. Trautwein, and J. Tischendorf, "Classification of colon polyps in NBI endoscopy using vascularization features," *Proc. SPIE*, vol. 7260, Feb. 2009, Art. no. 72602S.

[26] D. Mahapatra, P. Schueffler, J. A. W. Tielbeek, J. M. Buhmann, and F. M. Vos, "A supervised learning approach for Crohn's disease detection using higher-order image statistics and a novel shape asymmetry measure," *J. Digit. Imag.*, vol. 26, no. 5, pp. 920–931, Oct. 2013.

[27] Z. Wei, W. Zhang, J. Liu, S. Wang, J. Yao, and R. M. Summers, "Computer-aided detection of colitis on computed tomography using a visual codebook," in *Proc. IEEE 10th Int. Symp. Biomed. Imag.*, Apr. 2013, pp. 141–144.

[28] S. S. Ahmed, N. Dey, A. S. Ashour, D. Sifaki-Pistolla, D. Balas-Timar, V. E. Balas, and J. M. R. S. Tavares, "Effect of fuzzy partitioning in Crohn's disease classification: A neuro-fuzzy-based approach," *Med. Biol. Eng. Comput.*, vol. 55, no. 1, pp. 101–115, Jan. 2017.

[29] E. Mossotto, J. J. Ashton, T. Coelho, R. M. Beattie, B. D. MacArthur, and S. Ennis, "Classification of paediatric inflammatory bowel disease using machine learning," *Sci. Rep.*, vol. 7, no. 1, p. 2427, Dec. 2017.

[30] A. Alammari, A. R. Islam, J. Oh, W. Tavanapong, J. Wong, and P. C. de Groen, "Classification of ulcerative colitis severity in colonoscopy videos using CNN," in *Proc. 9th Int. Conf. Inf. Manage. Eng.*, 2017, pp. 139–144.

[31] G. Wimmer, A. Vécsei, M. Häfner, and A. Uhl, "Fisher encoding of convolutional neural network features for endoscopic image classification," *J. Med. Imag.*, vol. 5, no. 3, 2018, Art. no. 034504.

[32] S. Nadeem, M. A. Tahir, S. S. A. Naqvi, and M. Zaid, "Ensemble of texture and deep learning features for finding abnormalities in the gastro-intestinal tract," in *Proc. Int. Conf. Comput. Collective Intell.* Cham, Switzerland: Springer, 2018, pp. 469–478.

[33] G. Urban, P. Tripathi, T. Alkayali, M. Mittal, F. Jalali, W. Karnes, and P. Baldi, "Deep learning localizes and identifies polyps in real time with 96% accuracy in screening colonoscopy," *Gastroenterology*, vol. 155, no. 4, pp. 1069–1078, 2018.

[34] R. W. Stidham, W. Liu, S. Bishu, M. D. Rice, P. D. R. Higgins, J. Zhu, B. K. Nallamothu, and A. K. Waljee, "Performance of a deep learning model vs human reviewers in grading endoscopic disease severity of patients with ulcerative colitis," *JAMA Netw. Open*, vol. 2, no. 5, May 2019, Art. no. e193963.

[35] T. Ozawa, S. Ishihara, M. Fujishiro, H. Saito, Y. Kumagai, S. Shichijo, K. Aoyama, and T. Tada, "Novel computer-assisted diagnosis system for endoscopic disease activity in patients with ulcerative colitis," *Gastrointestinal Endoscopy*, vol. 89, no. 2, pp. 416–421, 2019.

[36] Y. Shin and I. Balasingham, "Comparison of hand-craft feature based SVM and CNN based deep learning framework for automatic polyp classification," in *Proc. 39th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2017, pp. 3277–3280.

[37] 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.

[38] D. Mahapatra, P. Schueffler, J. A. Tielbeek, J. M. Buhmann, and F. M. Vos, "A supervised learning based approach to detect Crohn's disease in abdominal MR volumes," in *Proc. Int. MICCAI Workshop Comput. Clin. Challenges Abdominal Imag.* Berlin, Germany: Springer, 2012, pp. 97–106.

[39] L. Han, M. Maciejewski, C. Brockel, W. Gordon, S. B. Snapper, J. R. Korzenik, L. Afzelius, and R. B. Altman, "A probabilistic pathway score (PROPS) for classification with applications to inflammatory bowel disease," *Bioinformatics*, vol. 34, no. 6, pp. 985–993, Mar. 2018.

[40] K. Pogorelov, K. R. Randel, C. Griwodz, S. L. Eskeland, T. de Lange, D. Johansen, C. Spampinato, D.-T. Dang-Nguyen, M. Lux, P. T. Schmidt, M. Riegler, and P. Halvorsen, "KVASIR: A multi-class image dataset for computer aided gastrointestinal disease detection," in *Proc. 8th ACM Multimedia Syst. Conf.*, Jun. 2017, pp. 164–169.

[41] Y. Maeda, S.-E. Kudo, Y. Mori, M. Misawa, N. Ogata, S. Sasanuma, K. Wakamura, M. Oda, K. Mori, and K. Ohtsuka, "Fully automated diagnostic system with artificial intelligence using endocytoscopy to identify the presence of histologic inflammation associated with ulcerative colitis (with video)," *Gastrointestinal Endoscopy*, vol. 89, no. 2, pp. 408–415, Feb. 2019.

[42] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[43] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," in *Proc. 31st AAAI Conf. Artif. Intell.*, 2017, pp. 4278–4284.

[44] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1251–1258.

[45] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4700–4708.
[46] B. Zoph, V. Vasudevan, J. Shlens, and Q. V. Le, "Learning transferable architectures for scalable image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8697–8710.
[47] F. Yu, V. Koltun, and T. Funkhouser, "Dilated residual networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 472–480.
[48] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.
[49] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, "Learning deep features for discriminative localization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 2921–2929.

SAHADEV POUDEL received the bachelor's degree in information technology from Purbanchal University, Nepal, in 2016, and the M.E. degree in IT convergence engineering from Gachon University, South Korea, in 2020. His current research interests include image classification, image matting, image segmentation, and deep learning.

YOON JAE KIM received the B.S., M.S., and Ph.D. degrees in medicine from the College of Medicine, Yonsei University, Seoul, South Korea, in 1999, 2003, and 2014, respectively. From March 2009 to March 2013, he was an Assistant Professor of gastroenterology with Gachon University Gil Medical Center, Incheon, South Korea. From April 2013 to March 2017, he was an Associate Professor and an Adjunct Professor with Gachon University Gil Medical Center, where he is currently a Professor. His current research interests include colon cancer, inflammatory bowel disease, medical device, and artificial intelligence for health care services.

DUC MY VO received the Bachelor of Engineering degree in automation from the University of Transport and Communication, Vietnam, in 2006, the master's degree from the Asian Institute of Technology, Thailand, in 2009, and the Ph.D. degree from the Laboratory of Prof. Dr. Andreas Zells, University of Tüebingen, Germany, in 2015. His thesis projects were focused on addressing the use of RGB-D images for six important tasks of mobile robots, such as face detection, face tracking, face pose estimation, face recognition, person detection, and person tracking. These topics have widely been researched in recent years, because they provide mobile robots with the abilities necessary to communicate with humans in natural ways. After one year of research with the Vietnam National Satellite Center, he joined the Computer Vision and Multimedia Laboratory, Chosun University, South Korea, in 2016. In 2017, he moved to the Software Department, Gachon University, south Korea. His current research interests include development of action recognition, face recognition, semantic segmentation, and biomedical imaging.

SANG-WOONG LEE (Senior Member, IEEE) received the B.S. degree in electronics and computer engineering and the M.S. and Ph.D. degrees in computer science and engineering from Korea University, Seoul, South Korea, in 1996, 2001, and 2006, respectively. From June 2006 to May 2007, he was a Visiting Scholar with the Robotics Institute, Carnegie Mellon University. From September 2007 to February 2017, he was a Professor with the Department of Computer Engineering, Chosun University, Gwangju, South Korea. He is currently a Professor with the Department of Software, Gachon University. His current research interests include face recognition, computational aesthetics, machine learning, and medical imaging analysis.