Biomedical Signal Processing and Control 68 (2021) 102654

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

A deep convolutional neural network for the detection of polyps in colonoscopy images

Tariq Rahim, Syed Ali Hassan, Soo Young Shin *

Kumoh National Institute of Technology, Gumi, Gyeongbuk 39177, Republic of Korea

ARTICLE INFO

Keywords:
Colonoscopy
Convolutional neural network
MISH
Polyp
Precision
Rectified linear unit
Sensitivity

ABSTRACT

Colonic polyps detection remains an unsolved issue because of the wide variation in the appearance, texture, color, size, and appearance of the multiple polyp-like imitators during the colonoscopy process. In this paper, a deep convolutional neural network (CNN) based model for the computerized detection of polyps within colonoscopy images is proposed. The proposed deep CNN model employs a unique way of adopting different convolutional kernels having different window sizes within the same hidden layer for deeper feature extraction. A lightweight model comprising 16 convolutional layers with 2 fully connected layers (FCN), and a Softmax layer as output layer is implemented. For achieving a deeper propagation of information, self-regularized smooth non-monotonicity, and to avoid saturation during training, MISH as an activation function is used in the first 15 layers followed by the rectified linear unit activation (ReLU) function. Moreover, a generalized intersection of the union (GIoU) approach is employed, overcoming issues such as scale invariance, rotation, and shape encountering with IoU. Data augmentation techniques such as photometric and geometric distortions are employed to overcome the scarcity of the data set of the colonic polyp. Detailed experimental results are provided that are bench-marked with the MICCAI 2015 challenge and other publicly available data set reflecting better performance in terms of precision, sensitivity, F1-score, F2-score, and Dice-coefficient, thus proving the efficacy of the proposed model.

1. Introduction

Early diagnosing of distinct diseases within the small intestine is a time-consuming and hectic process for physicians. This has led to the introduction of technologies such as colonoscopy and wireless capsule endoscopy [1] where several images are generated during those processes. Colorectal cancer (CRC) is the second-highest cause of death by cancer worldwide with 880,792 deaths and a mortality rate of 47.60% in 2018 reported by American Cancer Society [2]. A critical step in every CRC screening program is colonoscopy, where the aim is to classify and detect pre-malignant or malignant polyps employing a camera that is inserted into the large bowel [3]. 95.00% of CRC cases begin with the appearance of a growth on the inner lining of the rectum or colon, called a polyp. Various types of polyps exist including, adenoma polyps, which can worsen into CRC. CRC is curable in 90.00% of cases considering early detection [4]. Colonoscopy has emerged as a minimally invasive and additional tool for investigating polyps by examining the gastrointestinal tract [4]. The process of colonoscopy relies on highly experienced endoscopists, where recent clinical investigations have shown that the colonoscopy process misses 22.00–28.00% of polyps. These false negatives can lead to late diagnosis of colon cancer, resulting in a survival rate as low as 10.00% [5].

DL plays an important role in many areas, including text recognition tasks, self-driving cars, image recognition, and healthcare. Computer vision and ML-based techniques have evolved over several decades and are being employed for the automatic detection of polyps [6–8]. The candidate boundaries features of polyps are obtained using low-level image processing techniques such as hand-crafted features like texture, histograms of oriented gradients, color wavelets, and local binary patterns analysis [9,10]. Recently, advanced algorithms have been suggested to evaluate polyp appearance based on factors such as context information [11] and edge shape [12]. However, the decrease in detection performance is mainly due to the similar appearance of polyp-like and polyp structures that need to be addressed.

Convolutional neural networks (CNNs) display promising results when it comes to object detection and segmentation. In the 2015 MICCAI challenge, CNN outperformed techniques based on low image processing, i.e. hand-crafted features analysis for the detection of polyp [6].

* Corresponding author.
E-mail addresses: tariqrahim@ieee.org (T. Rahim), Syedali@kumoh.ac.kr (S.A. Hassan), wdragon@kumoh.ac.kr (S.Y. Shin).

https://doi.org/10.1016/j.bspc.2021.102654
Received 1 October 2020; Received in revised form 31 March 2021; Accepted 16 April 2021
Available online 28 April 2021

In the last decade, the region-based CNN approaches, such as *R-CNN* [13], *Fast R-CNN* [14], and *Faster R-CNN* [15] presented a promising results for objects and polyps detection. Regression-based attempts are performed by employing a single-shot multi-box detector (SSD) [16] and You Only Look Once (YOLO) [17] for the detection of polyps. Despite CNN's robustness and high detection efficiency, recent investigations have shown that deep neural networks (DNNs), including CNN's, are extremely vulnerable to noise up to one single-pixel [18] and perturbations [19] can lead to miss detection. Even though computer-aided detection techniques can effectively classify and detect, polyps detection remains challenging due to its significant size, appearance, and intensity variations within the small bowel and consecutive frames. Moreover, the artifacts generated due to the existence of intestinal content (fecal and bubble particles), as well as the appearance of specular overexposed areas and highlights, may extra aggravate the circumstance. This is a serious issue, because polyps and polyp-like objects have similar appearances in consecutive frames, leading to miss-detection even when implementing powerful models such as CNN. Furthermore, the performance of DL approaches is profoundly associated with the amount of data available for training. The lack of availability of labeled polyp images for training makes the detection and segmentation of the polyp a difficult task [20].

Considering the above critical issues related to a DL-based model for the detection of the polyp, this work presents a deep CNN-based detection model of polyp in colonoscopy images. The proposed deep CNN model employs a unique way of adopting different convolutional kernels having different window sizes within the same hidden layer for deeper feature extraction. A lightweight model comprising 16 convolutional layers with 2 fully connected layers (FCN), and a Softmax layer as output layer is implemented. For achieving a deeper propagation of information, self-regularized smooth non-monotonicity, and to avoid saturation during training, MISH [21] as an activation function is used in the first 15 layers followed by the rectified linear unit activation (ReLU) function. Moreover, a generalized intersection of the union (GIoU) [22] approach is employed, overcoming issues such as scale invariance, rotation, and shape encountering with IoU. The rest of the paper is categorized as follows: Section 2 presents recent related work done for polyp detection in colonoscopy images. Section 3 details the proposed deep CNN model for polyp detection in colonoscopy images while in Section 4, the experimental results are described in detail, along with the data set specifications and data augmentation process. Finally, in Section 5 the paper is wrapped with a conclusion with the future work presentation.

2. Related work

From the past two decades, techniques based on computer vision and machine learning (ML) have been introduced for the computerized detection of polyps [19]. In initial investigations, hand-craft features, such as texture, Haar, color wavelet, the histogram of oriented gradients (HoG), and local binary pattern (LBP) were studied for the detection of polyps [23–25]. More advanced algorithms were introduced; where edge shape and context information was used in the former while valley information based on polyp representation was practiced in the latter. These feature patterns are commonly alike in polyp-like normal and polyp structures, leading to a decreased performance [19].

As CNN techniques for single and multiple object detection advance, they increasingly outperform previous conventional image processing techniques [26]. A two-stage detection of a polyp where the region of interest (RoI) is detected using CNN followed by a false positive (FP) reduction unit is proposed [19]. For the detection and segmentation, some of the state-of-the-art techniques are bench-marked investigating each method's efficacy and speed [27]. For multiple object detection, a region-based CNN combined with a deformable part-based model has been proposed to handle feature extraction and occlusion [28]. Recently, with the progress of DL in multiple image processing applications, a CNN-based method has been introduced for polyp detection [29,30]. Due to this and related progress, CNN features outperformed hand-crafted features in the MICCAI 2015 polyp detection challenge [6]. A few-shot anomaly of polyp classification problem is presented that successfully recognize, localize, and classify using an imbalanced data set [3] A regression-based CNN model using ResYOLO combined with efficient convolution operators has been shown to successfully track and detect polyps in colonoscopy videos [7]. Automatic detection of hyperplastic and adenomatous colorectal polyps in colonoscopy images has been performed using sequentially connected encoder–decoder-based CNN [31]. A unique U-Net is proposed that is comprised of fully 3D layers enabling the network to be fed with hyperspectral images and an output layer of Dice prediction for probabilistic approaches [32]. Furthermore, automatic polyp detection in colonoscopy videos can be conducted via ensemble CNN, which learns a variety of polyp features such as texture, color, shape, and temporal information [33].

To surmount the scarcity of adequate training samples for the use of pre-trained CNNs on large-scale natural images, transfer learning systems have been proposed. This has been successfully implemented in several medical applications, such as automatic interleaving between radiology reports and diagnostic CT [34], MRI imaging, and ultrasound imaging [35]. Moreover, the performance of transfer learning-based CNN models such as AlexNet and GoogLeNet has been evaluated for classification of intestinal lung diseases and detection of thoracic-abnormal lymph nodes [36]. Similarly, a transfer learning-based method using the deep-CNN model Inception Resnet has been used to detect polyps in colonoscopy images [37]. Questions of whether a CNN with adequate fine-tuning can overcome the full training of the model from scratch have been answered in detail by examination of four different medical imaging applications in three different specialties: gastroenterology, radiology, and cardiology for the purpose of classification, detection, and segmentation [38].

CNN has been used for decades in the field of computer vision and ML for various applications. However, training a deep CNN model from scratch is a complicated task [38]. Deep CNN models require a large amount of labeled training data and become a tough demand when large-scale annotated medical data sets are unavailable. Training the models is tedious and computationally time-consuming; becomes worse even more so when facing complications such as over-fitting and convergence issues. To overcome these issues, this work presents a deep CNN model for the detection of polyps in colonoscopy images. The proposed model employs fewer hidden layers, making the model lighter and less time-consuming during training. In general, the main contributions of the paper are as follows:

- A lightweight DL-based model comprising 16 convolutional layers with 2 fully connected layers (FCN), and a Softmax layer as output layer is implemented for the detection of polyp within colonoscopy images.
- The proposed model employs two activation function, i.e. *Mish* in the first 15 layers for deeper propagation of information, self-regularized smooth non-monotonicity, and to avoid saturation during training followed by *ReLU* function for gradient flow.
- GIoU approach is employed, overcoming issues such as scale invariance, rotation, and shape encountering with IoU. This approach minimize the overlapping issue of bounding box when detecting polyp and polyp like structure.
- To evaluate the efficiency of the proposed model, a detailed experimental results are provided that are bench-marked with the MICCAI 2015 challenge and other publicly available data set.

3. Proposed deep convolutional neural network (CNN) architecture

At first, the input image is divided into grid cells $G \times G$ during the

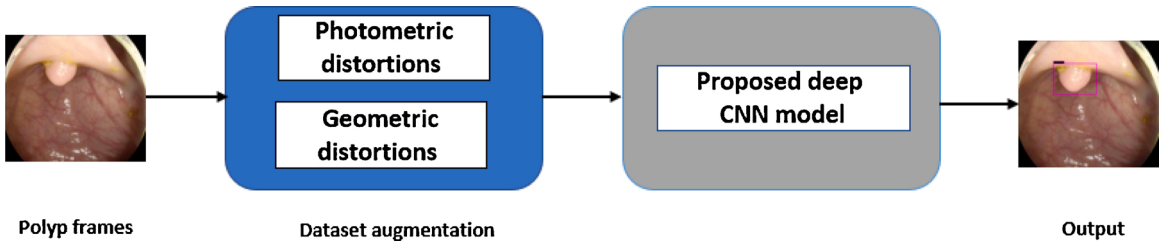

Fig. 1. Flow of the proposed deep CNN model for polyp detection.

Fig. 2. Architecture of the proposed deep CNN for polyp detection.

training phase. The size of the grid cells for the proposed implementation is kept as 8×8 resulting in 64 cells, where each grid cell is responsible for the probability of the presence of the object within the image. Each grid cell is associated with a distinct region of the image, and these cells predict objects whose centers lay in the region. This allows the network to have a structured output description to utilize the advantage of spatial regularity of any images. Then the image is labeled using the RectLabel tool, generating a bounding box "B" consisting of five predictions. Here each grid cell predicts "B" that represents the rectangle that encompasses an object and confidence score "C_s" for each class of the object within. These predictions include the horizontal and vertical components reflecting the mid-point of the image within the grid cell and are labeled as "x", and "y", respectively. The height and width are labeled as "h" and "w" representing the height and width of the "B", respectively. Finally, a confidence score "C_s" is defined for each defined grid cell that actually reflect the certainty or how confident the model is that the box comprises an object. The "C_s" also represent how precise it thinks the box is that predicts. The objective function of bounding box "B" is a bag of freebies using mean square error (MSE) to perform regression on the center coordinate points, height, and width of the box "B". The intersection over union (IoU) also named as Jaccard index is a vital indicator for estimating the distance between the predicted truth and the ground truth "B" and used as a metric to compare the similarity among two arbitrary shapes [39]. The IoU helps to find the object by computing that the amount of overlap occurring between predicted truth and the ground truth "B" [17]. The IoU is can be formulate in a generalized form as:

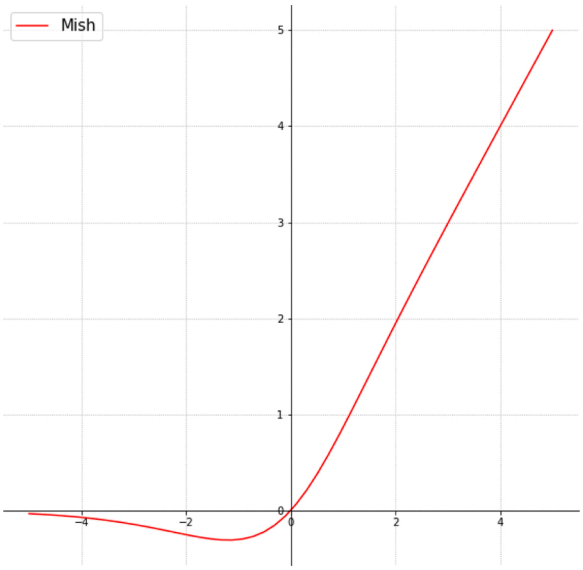

Fig. 3. Mish activation function.

$$\mathrm{I\,o}U = \frac{|E \cap F|}{|E \cup F|} = \frac{|I|}{|U|} \tag{1}$$

where "E" and "F" represent the predicted truth and ground truth, respectively. Here, the IoU distance $\mathcal{L}_{\mathrm{I\,o}U} = 1 - \mathrm{I\,o}U$ [39] fulfills all properties of a metric, including the identity of indiscernibles, non-negativity, triangle inequality, and symmetry, but has a scale-invariant issue [22]. The cost or loss function for object detection use l_1 norm and l_2 norm for x, y, w, h, but due to the scale-invariant property of IoU, there is an increase in loss with respect to scaling. In the proposed deep CNN approach to polyp detection, we have implemented generalized (GIoU) [22] as a new loss to optimize the non-overlapping "B" in consideration of the shape and orientation of the object in "B". The GIoU finds the smallest convex shape $C \subseteq \mathbb{S} \in \mathbb{R}^n$ for two arbitrary convex shapes $E, F \subseteq \mathbb{S} \in \mathbb{R}^n$ followed by the calculation of the ratio between the area occupied by C minus "E" and "F", divided by the total area occupied by "C". Details for the algorithm and formulation can be found in [22], where a GIoU is expressed in simple mathematical form as, $\mathrm{G\,Io\,U} = \mathrm{I\,o}U - \frac{|C \setminus (E \cup F)|}{|C|}$. Furthermore, we have applied a non-maximum suppression algorithm involving "C_s" to avoid multiple and overlapping GIoUs.

Fig. 1 shows the general flow of the proposed approach for polyp detection in colonoscopy images. As shown in Fig. 2, the proposed deep CNN consists of 16 convolutional layers, two fully connected layers, and a softmax layer. The number of convolutional layers is finalized considering the training time and generalization trade-off. This was done after extensive hit and trial to minimize the computational complexity to obtain the smallest prediction error and simpler model or not compromising the accuracy. Considerations such as placement of two activation functions after setting the number of convolutional layers are performed after hit and trail while analyzing the training time and final detection output precision. To lessen computational complexity and improve hierarchical image features, *maxpooling* is used for the first 15 convolutional layers. For better image feature extraction, different sizes of *convolutional filers* such as 16, 32, 64, 128, 256, 512, and 1026 with *kernel size* of 3×3 are employed. The choice of different size of *convolutional filters* in each layer is based on the concept that some of the important features are local and can be found in more than one spatial position within the image. So, for deeper feature extraction, and to avoid not missing any useful feature, the convolutional layers operate with different *convolutional filters*. In the proposed model, we have implemented *Mish* [21], which is a self-regularized smooth non-monotonic activation function, in the first 15 convolutional layers. This implementation was done after extensive trials to find the best matching position of the activation function. As observed in Fig. 3, *Mish* is an unbounded above result in avoiding saturation due to capping. This may normally lead to slow training, i.e., near-zero gradients. A better gradient flow and smooth propagation of information across deeper layers are achieved by the infinite order of continuity and a small allowance of negative values, in comparison to a strictly bounded rectified linear unit (ReLU) as an activation function. MISH can be expressed mathematically as:

$$f(x) = x \cdot tanh(\zeta(x)) \tag{2}$$

where $\zeta(x) = ln(1 + e^x)$ is softplus activation [21].

In the last layers, *ReLU* is used as an activation function to reduce the likelihood of gradient vanishing and achieve the sparsity. The strategy of using two activation functions *Mish* and *ReLU* results in smooth propagation of information across deeper layers. *MISH* helps to avoid capping, and *ReLU* prevents the gradient from vanishing. So *Mish* which is unbounded above is used in the first 15 layers avoids the saturation that can lead to slow training, while its unbounded below property results in a vigorous regularization effect [21] as shown in Fig. 3. In the last layer, *ReLU* is used to achieve sparsity and stochastic gradient descent (SGD) convergence acceleration. Flattening is done by two fully connected layers to yield a single continuous linear vector followed by *softmax* or the regression layer to generate the required output. Softmax designates decimal probabilities per class in a multi-class issue which should add up to 1.0 as a whole and act to predict a multinomial probability distribution. This supplementary constraint improves training converge more speedily. Furthermore, the purpose why Softmax is beneficial is because it transforms the output of the last layer in the model into what is a probability distribution and also resilient to outliers [40].

The proposed deep CNN is trained with a similar concept of multilayer perceptions, i.e., a back-propagation algorithm which minimizes the cost function concerning the unknown weights "W" [41]:

$$\mathcal{L} = -\frac{1}{|L|}\sum_{i}^{|L|} ln\left(p\left(m^i|L^i\right)\right) \tag{3}$$

where $|L|$ represents the number of training images, $p(m^i|L^i)$ represents the probability that L^i is accurately classified, and L^i represents the ith training image with the associated label m^i. We have applied SGD as an optimizer, which minimizes the cost function over the whole training data set along with the cost over mini-batches of data. If W_j^t represents the weights in the jth convolutional layer at t iteration, and $\widehat{\mathcal{L}}$ represents the cost over a mini-batch of size M, then in the next iteration the updated weights are calculated as given below:

$$\begin{aligned} \gamma^t &= \gamma^{\lfloor t\ M/|X| \rfloor} \\ V_j^{t+1} &= \mu V_j^t - \gamma^t \eta_j \frac{\partial \widehat{\mathcal{L}}}{\partial W_j} \\ W_j^{t+1} &= W_j^t + V_j^{t+1} \end{aligned} \tag{4}$$

where η_j is the learning rate of the jth, μ is the momentum indicating the previously updated weight contribution in the current iteration, and γ represent the scheduling rate which after each epoch decreases the learning rate η. In Eq. (4), V_j^t is the learnable parameter in the jth convolutional layer at t iteration while V_j^{t+1} represent the next learnable parameter at $t+1$ iteration [42]. The simulation parameters used for the proposed deep CNN are given in Table 1.

Table 1
Simulation parameters used for deep CNN model.

Network parameters	Configuration values
Image size	448×448
Learning rate (η)	0.0001
Momentum	0.9
Batch size	32
Iterations (t)	10,000

4. Experimental results and discussion

This section details the data set specifications and experimental results generated by implementing the proposed deep CNN for the detection of polyps in colonoscopy images.

4.1. Dataset specifications and augmentation

The study used a publicly available data set of polyp-frames obtained from the ETIS-LARIB database [43], containing 196 polyp images. These images were obtained from 34 different colonoscopy videos of 44 different polyps with various appearances and sizes, having a resolution of 1225×966 pixels. The ground truth of polyp areas for polyp data sets is determined by expert video endoscopists. As the acquired data set contained 196 polyp images and considering vivid variations of the polyp structure, the CNN model trained with such a small amount of data is likely to be meaningless and unstable. Therefore, Data augmentation is performed on the colonoscopy images to avoid over-fitting. In colonoscopy imagery, polyps exhibit large variations in location, color, and scale. Moreover, variations in brightness and definition also occur due to varying the view-point of the camera. Therefore, in addition to photometric distortions and geometric distortions, zooming, shearing, and altering brightness as strategies for data augmentation are adopted.

For photometric distortions, we controlled brightness and contrast as an enhancement, while blurring by adding noise with a standard deviation (σ) of 1.0. Similarly, for geometric distortions, clock-wise rotation of the polyp images with angles of 90°, 180°, and 270° were performed. Zoom-in and zoom-out with zooming parameters such as 30.00% and 10.00% were performed to obtain different scales of polyp images. Lastly, shearing for both the x-axis and the y-axis was performed to shear the images from left to right and top to bottom, respectively. Fig. 4 shows photometric and geometric forms of image augmentation. In this way, we augmented the data set of the ETIS-LARIB database from 196 polyp images to 2156 images, which is more suitable for training the proposed deep CNN model.

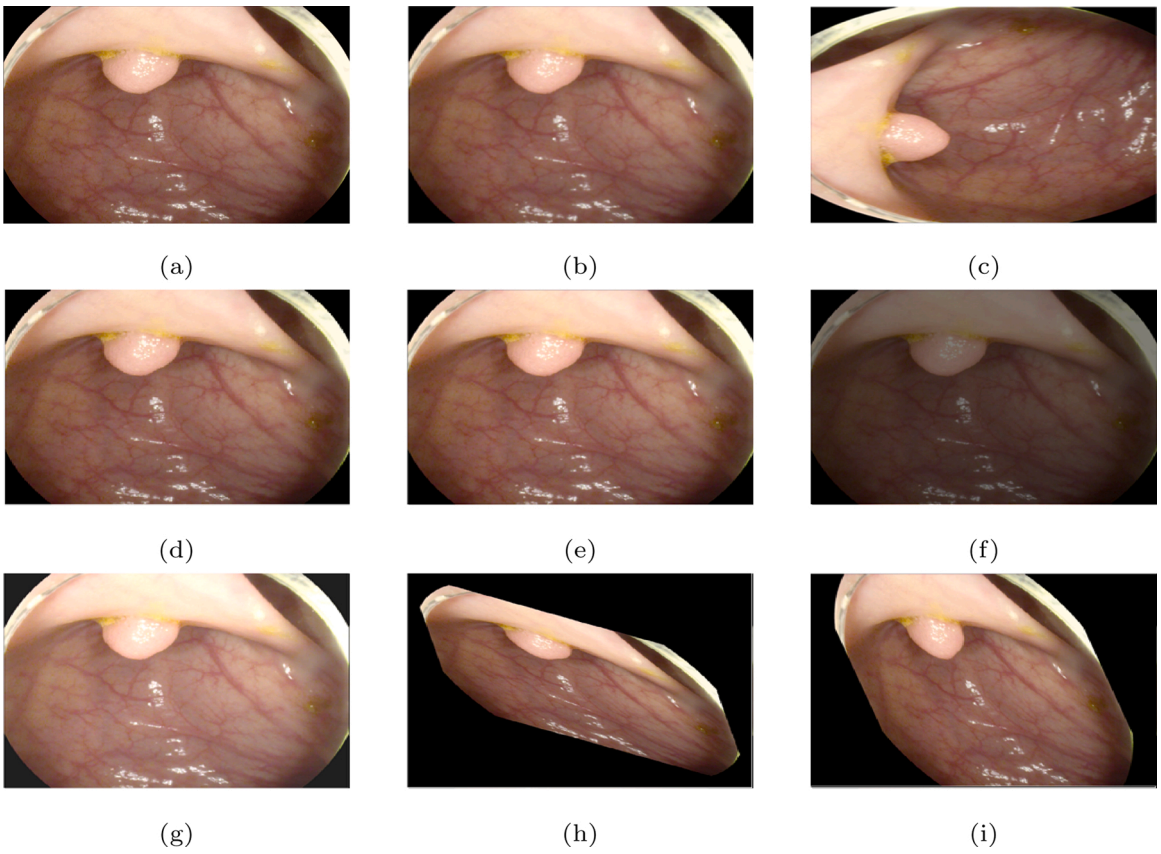

Fig. 4. Image from the dataset with photometric and geometric augmentation. (a) Original frame of polyp, (b) noisy polyp frame with $\sigma = 1.1$, (c) rotated polyp frame with 90°, (d) polyp frame with 15.00% zoom in, (e) polyp frame with 15.00% zoom out, (f) dark polyp frame, (g) bright polyp frame, (h) sheared polyp frame by y-axis, (i) sheared polyp frame by x-axis.

4.2. Performance metrics for evaluations

The metrics used to evaluate the detection of polyps within colonoscopy frames in this work are the same as those used in the MICCAI 2015 challenge [6]. The output obtained using the proposed model has rectangular shaped coordinates (x,y,w,h). The following parameters are defined as follows:

True Positive (TP): True output detection if the detected centroid falls within the polyp ground truth. For multiple true output detection within the same frame and of the same polyp, TP is counted as one.

True Negative (TN): True detection, i.e., negative frames (frames without polyps) yielding no detection output.

False Positive (FP): False detection output where the detected centroid falls outside the polyp ground truth.

False Negative (FN): False detection output, i.e, polyp is missed in a frame having a polyp.

Employing the above parameters, we can compute the following performance metrics to efficiently evaluate the performance of the proposed deep CNN model.

Precision: This metric computes how precisely the model is detecting a polyp within an image

$$\text{Precision} \quad (\text{Pre}) = \frac{\text{T P}}{\text{T P} + \text{F P}} \times 100 \tag{5}$$

Sensitivity: This metric is also called recall or True Positive rate and computes the proportion of the actual polyps that were detected correctly

$$\text{Sensitivity} \quad (\text{Sen}) = \frac{\text{T P}}{\text{T P} + \text{F N}} \times 100 \tag{6}$$

F1-score and F2-score: F1 and F2-score is simply the harmonic mean between precision and sensitivity, in a range of $[0,1]$. Both scores are recognized to balance the precision and sensitivity. The F1-score is given as:

$$F1\text{-score} = \frac{2 \times \text{Sen} \times \text{Pre}}{\text{Sen} + \text{Pre}} \times 100 \tag{7}$$

while the F2-score can be calculated as:

$$F2\text{-score} = \frac{5 \times \text{Pre} \times \text{Sen}}{4 \times \text{Pre} + \text{Sen}} \tag{8}$$

Dice Coefficient: This metric is used for pixel-wise result comparison between ground truth and predicted detection that ranges $[0,1]$, and is given as:

$$\text{Dice} \quad \text{coefficient}\,(E,F) = \frac{2 \times |E \cap F|}{|E| + |F|} = \frac{2 \times \text{T P}}{2 \times \text{T P} + \text{F P} + \text{F N}} \tag{9}$$

4.3. Polyp frames evaluation

This section reports the polyp detection performance of the proposed CNN model. For implementation of the model, 80.00% and 20.00% of the 2156 augmented polyp frames were used for training and testing, respectively. Fig. 5 shows the real-time training phase of the proposed model, where 10,000 iterations were run to achieve the best weights. The model was trained using the simulation parameters as given in Table 1, both for the non-augmented ETIS-LARIB database [43] containing 196 poly images, and the augmented data set, for fair performance comparison. A high mean average precision of 97.70% with an MSE of 0.900 was obtained in the early iterations, resulting in the best weights for testing purposes.

Table 2 shows the performance of the proposed deep CNN model for

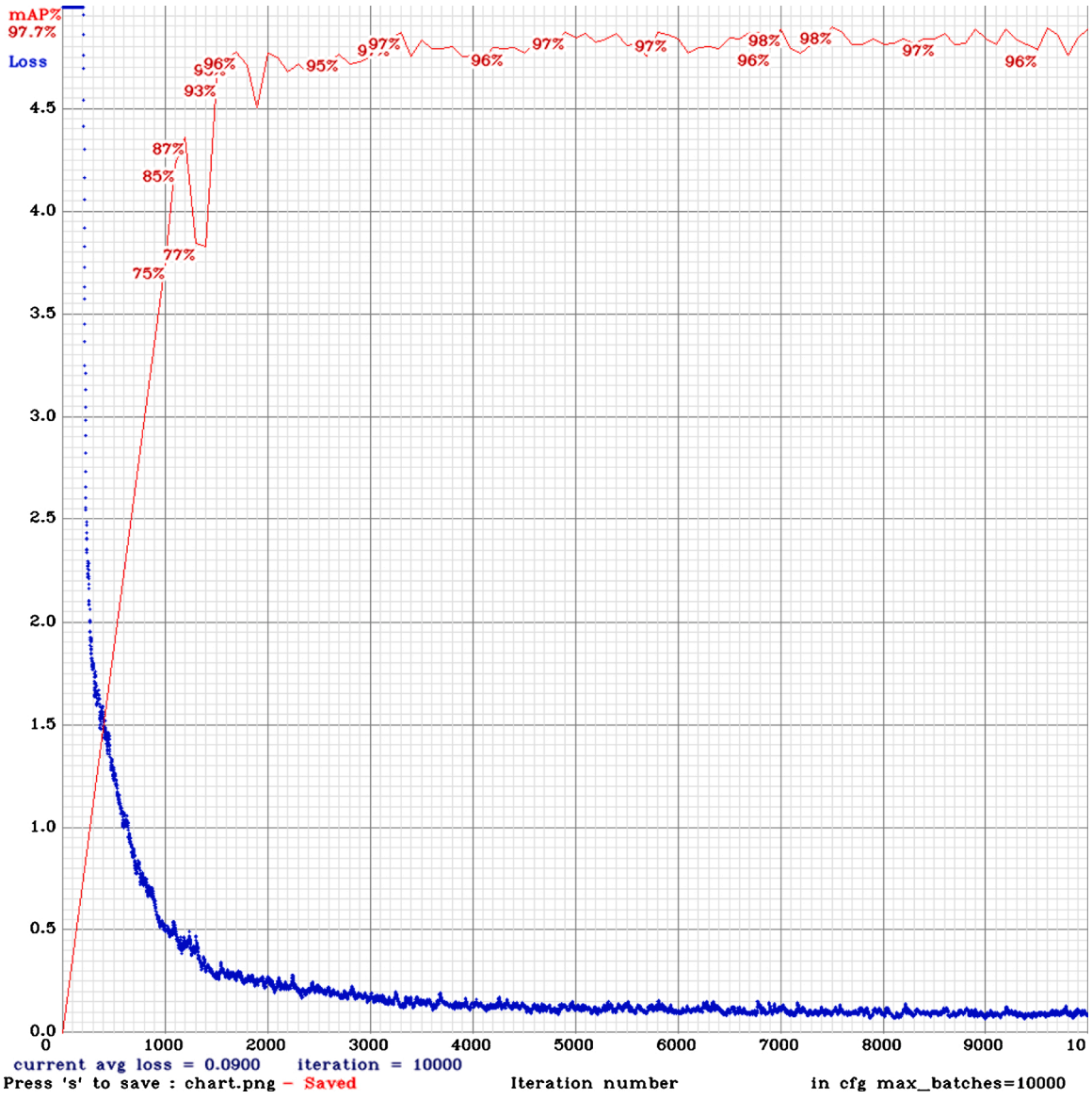

Fig. 5. Training phase of the proposed deep CNN model.

Table 2
Detection performance comparison of the proposed deep CNN model on ETIS-Larib database without(w/0) and with augmentation strategies.

Data set	Performance metrics			
	[37] (%)		Propose deep CNN model (%)	
Non-augmented ETIS LARIB database (196)	Pre	48.00	Pre	72.00
	Sen	39.40	Sen	63.82
	F1-score	43.30	F1-score	67.66
	F2-score	40.90	F2-score	65.30
	Dice-coefficient	NA	Dice-coefficient	0.676
Augmented ETIS LARIB database (2156)	Pre	91.40	Pre	94.44
	Sen	71.20	Sen	82.92
	F1-score	80.00	F1-score	88.30
	F2-score	74.50	F2-score	85.00
	Dice-coefficient	NA	Dice-coefficient	0.88

the detection of polyp using the performance metrics describes in Section 4.2. It can be observed that the proposed deep CNN model outperformed the bench-marked work both for non-augmented and augmented data sets. The bench-marked method implements a pre-trained model of (Inception Resnet) while employing a region proposal network (RPN). The RPN model acts a detection model at post-learning stage [37]. Furthermore, the high-performance values of precision, the F1-score, the F2-score, and the Dice coefficients with low sensitivity performance value reflect the stability and efficacy of the proposed model. For the non-augmented data set of ETIS-LARIB, the generated TP, FP, and FN values were 90, 35, and 51, respectively. Similarly, 20.00% of the augmented data set was employed for testing purposes, generating TP, FP, and FN values of 340, 20, and 70, respectively.

The results shown in Fig. 6 are generated using the proposed deep CNN model on the augmented data set. It can be observed that the proposed model shows better polyp detection performance. As illustrated in Fig. 6, polyps within a frame can be identified at multiple positions, and as noted above in this case, the TP for detection is considered to be 1. The proposed deep CNN model performed better than other benchmark results in terms of the performance metrics listed above, as shown in Table 2 and Fig. 6.

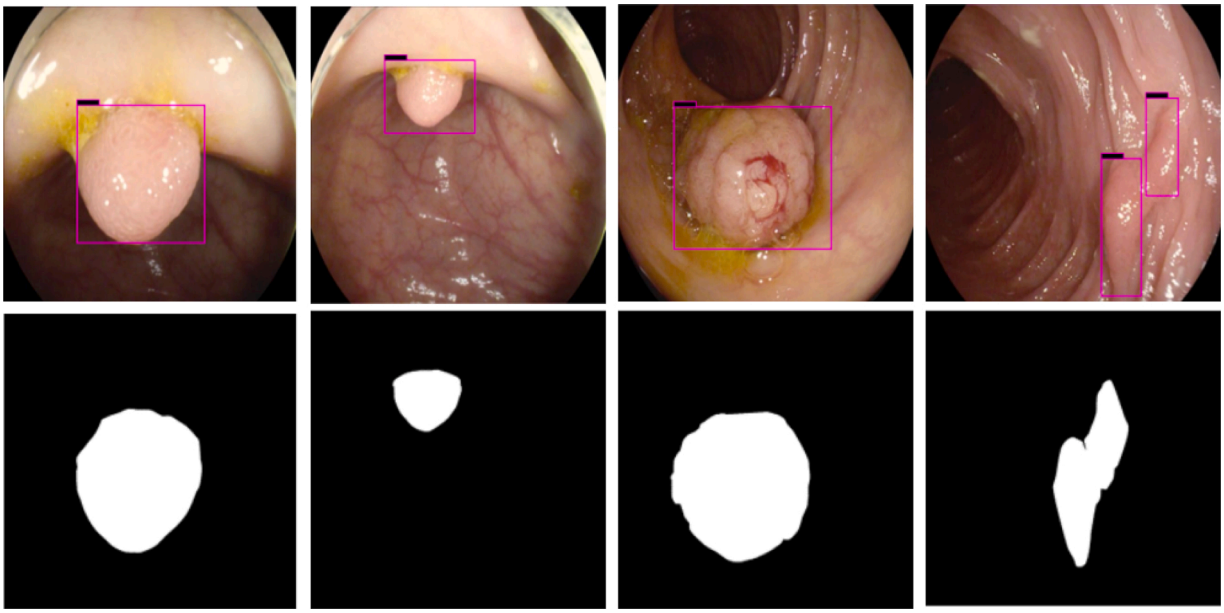

Fig. 6. Example of accurate detection along with the correct ground truth using deep CNN model. The first row shows the detection results for different polyp from data augmentation process. The second shows the ground truth images of test images.

Table 3
Simulation parameters used for bench-marked approach [44].

Network parameters	Configuration values
Image size	448 × 448
Learning rate (η)	0.001
Momentum	0.9
Batch size	16
Epochs	50

In order to evaluate the efficacy of the proposed deep CNN model for the detection of polyp, a publicly available database, i.e., CVC-ClinicDB database is used that comprised of 612 polyp images [44]. For fair comparison with ETIS LARIB augmented data set, the CVC-ClinicDB data set is also augmented resulting in 2156 images. This comparison was done by using both the data set opted, i.e., CVC-ClinicDB and the simulation parameters [44]. Table 4 shows the bench-marked results on publicly available data sets; clearly it can be observed that proposed deep CNN model reflect a high performance for the detection of polyp. The simulation parameters for the proposed deep CNN model and bench-marked approach, i.e., YOLOv2 on public data set are given in Tables 1 and 3, respectively.

For single and deep layer of the proposed model, we have shown channel activation representing the convolutional kernels accurately detected the polyp. Fig. 7 shows different bright and dark parts corresponding to the spatial property of the object within the test images for single and deep layers. The *top left* is the test polyp image followed by *top right* detection output generated by proposed deep CNN model. The *bottom left* shows the single layer activation channel whole *bottom right* shows the deep layer for deeper feature analysis represented by green rectangular boxes. It can be observed in Fig. 7, that both single and deep layers are extracting polyp features with a high score, resulting in high polyp detection (Table 4).

4.4. Benchmark performance with other approaches

To evaluate the efficacy of the proposed deep CNN model, the results from the 2015 MICCAI challenge [6] is bench-marked, as the data sets used were the same. Among the challenge, the top three experimental results from each team in the challenge, i.e., UNS-UCLAN, OUS, and CUMED were selected for bench-marking. The reason was that CNN has been employed for learning end-to-end detection of the polyp. The UNS-UCLAN team [6] used three CNNs for the extraction of features on multiple spatial scales, followed by a classification approach with a multi-layer perception network. AlexNet, a CNN-based model was adopted with a conventional sliding window to perform patch-based classification [37]. The CUMED team used a segmentation approach based on CNN [43], where classification was conducted pixel-wise along with a ground truth mask.

As shown in Table 5, the generated results from the proposed model using the augmented data set outperform the other team's methods on several metrics, including precision, sensitivity, F1-score, F2-score, and Dice-coefficient. As DL-based methods employ different computers with different specifications, it is hard to benchmark detection processing directly. In our work, the data set was trained and tested on NVIDIA Titan RTX GPUs to reduce processing time. Compared to the other studies listed in Table 2 the mean detection processing time 0.6 s per frame. This is slightly greater than that in competing models, but the increased processing time comes with better performance.

5. Conclusion

In this paper, we presented a computerized DL-based detection model for colonic polyps. a deep convolutional neural network (CNN) based model for the computerized detection of polyps within colonoscopy images is proposed. The proposed deep CNN model employs a unique way of adopting different convolutional kernels having different window sizes within the same hidden layer for deeper feature extraction. A lightweight model comprising 16 convolutional layers with 2 fully connected layers (FCN), and a Softmax layer as output layer is implemented. For achieving a deeper propagation of information, self-regularized smooth non-monotonicity, and to avoid saturation during training, MISH as an activation function is used in the first 15 layers followed by the rectified linear unit activation (ReLU) function. Moreover, a generalized intersection of the union (GIoU) approach is employed, overcoming issues such as scale invariance, rotation, and

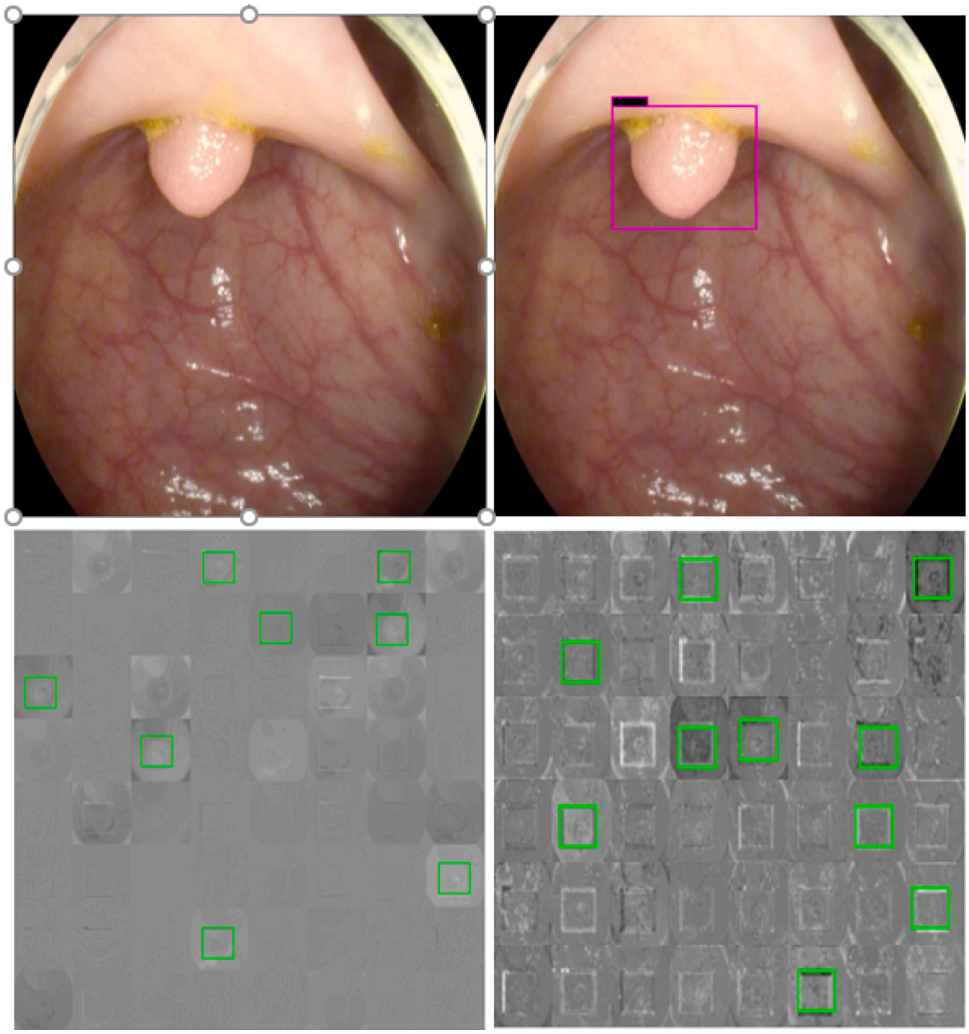

Fig. 7. Test polyp channel activation visualization of CNN after training of the proposed deep CNN model. Top left: test polyp images, Top right: detection output of the test image, Bottom left: activation channel for single layer, Bottom right: activation channel for deep layer (both shown by green rectangular box).

Table 4
Detection performance comparison of the proposed deep CNN model on ETIS-LARIB and CVC-ClinicDB databases.

ETIS LARIB database (Proposed deep CNN approach) 2156 images	Performance metrics						CVC-ClinicDB (Proposed deep CNN approach) 2156 images	Performance metrics					
	Pre	Sen	F1-score	F2-score	Dice-coefficient	AUC		Pre	Sen	F1-score	F2-score	Dice-coefficient	AUC
	94.44 (%)	82.92 (%)	88.30 (%)	85.00 (%)	0.88	90.42 (%)		94.01 (%)	82.00 (%)	87.87 (%)	84.24 (%)	0.80	89.02 (%)
							Bench-marking with [44]	NO	90.42 (%)	NO	NO	NO	92.33 (%)

shape encountering with IoU. Data augmentation techniques such as photometric and geometric distortions are employed to overcome the scarcity of the data set of the colonic polyp. Detailed experimental results are provided that are bench-marked with the MICCAI 2015 challenge and other publicly available data set reflecting better performance in terms of precision, sensitivity, F1-score, F2-score, and Dice-coefficient, thus proving the efficacy of the proposed model.

Table 5
Performance comparison of the proposed deep CNN model on ETIS-Larib database with other methods.

Implemented methods	Performance metrics				
	Pre (%)	Sen (%)	F1-score (%)	F2-score (%)	Dice-coefficient
UNS-ULCAN	32.70	52.80	40.40	47.10	NA
OUS	69.70	63.00	66.10	64.20	NA
CUMED	72.30	69.20	70.70	69.80	NA
Proposed deep CNN model	94.44	82.92	88.30	85.00	0.88

Acknowledgments

This work was supported by Priority Research Centers Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology" (2018R1A6A1A03024003). This research was supported by the MSIT (Ministry of Science and ICT), Korea, under the Grand Information Technology Research Center support program (IITP-2021-2020-0-01612) supervised by the IITP (Institute for Information & communications Technology Planning & Evaluation).

Conflict of interest: None declared.

References

[1] T. Rahim, M.A. Usman, S.Y. Shin, A Survey on Contemporary Computer-Aided Tumor, Polyp, and Ulcer Detection Methods in Wireless Capsule Endoscopy Imaging, 2019. arXiv:1910.00265.
[2] R. Segal, K. Miller, A. Jemal, Cancer statistics, 2018, CA Cancer J. Clin. 68 (1) (2018) 7–30.
[3] Y. Tian, L.Z.C.T. Pu, Y. Liu, G. Maicas, J.W. Verjans, A.D. Burt, S.H. Shin, R. Singh, G. Carneiro, Detecting, Localising and Classifying Polyps From Colonoscopy Videos Using Deep Learning, 2021. arXiv:2101.03285.
[4] O. Chuquimia, A. Pinna, X. Dray, B. Granado, Polyp follow-up in an intelligent wireless capsule endoscopy, in: 2019 IEEE Biomedical Circuits and Systems Conference (BioCAS), IEEE, 2019, pp. 1–4.
[5] A. Leufkens, M. Van Oijen, F. Vleggaar, P. Siersema, Factors influencing the miss rate of polyps in a back-to-back colonoscopy study, Endoscopy 44 (05) (2012) 470–475.
[6] J. Bernal, N. Tajkbaksh, F.J. Sánchez, B.J. Matuszewski, H. Chen, L. Yu, Q. Angermann, O. Romain, B. Rustad, I. Balasingham, et al., Comparative validation of polyp detection methods in video colonoscopy: results from the MICCAI 2015 endoscopic vision challenge, IEEE Trans. Med. Imaging 36 (6) (2017) 1231–1249.
[7] R. Zhang, Y. Zheng, C.C. Poon, D. Shen, J.Y. Lau, Polyp detection during colonoscopy using a regression-based convolutional neural network with a tracker, Pattern Recogn. 83 (2018) 209–219.
[8] N. Tajbakhsh, S.R. Gurudu, J. Liang, System and methods for automatic polyp detection using convulutional neural networks, US Patent 10,055,843 (2018).
[9] W. Wang, J. Tian, C. Zhang, Y. Luo, X. Wang, J. Li, An improved deep learning approach and its applications on colonic polyp images detection, BMC Med. Imaging 20 (1) (2020) 1–14.
[10] J. Bernal, J. Sánchez, F. Vilarino, Towards automatic polyp detection with a polyp appearance model, Pattern Recogn. 45 (9) (2012) 3166–3182.
[11] N. Tajbakhsh, S.R. Gurudu, J. Liang, Automated polyp detection in colonoscopy videos using shape and context information, IEEE Trans. Med. Imaging 35 (2) (2015) 630–644.
[12] J. Bernal, F.J. Sánchez, G. Fernández-Esparrach, D. Gil, C. Rodríguez, F. Vilariño, Wm-dova maps for accurate polyp highlighting in colonoscopy: validation vs. saliency maps from physicians, Comput. Med. Imaging Graph. 43 (2015) 99–111.
[13] R. Girshick, J. Donahue, T. Darrell, J. Malik, Rich feature hierarchies for accurate object detection and semantic segmentation, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2014) 580–587.
[14] R. Girshick, Fast r-cnn, Proceedings of the IEEE International Conference on Computer Vision (2015) 1440–1448.
[15] S. Ren, K. He, R. Girshick, J. Sun, Faster r-cnn: towards real-time object detection with region proposal networks, IEEE Trans. Pattern Anal. Mach. Intell. 39 (6) (2016) 1137–1149.
[16] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, A.C. Berg, Ssd: single shot multibox detector, in: European Conference on Computer Vision, Springer, 2016, pp. 21–37.
[17] J. Redmon, A. Farhadi, Yolov3: An Incremental Improvement, 2018. arXiv: 1804.02767.
[18] J. Su, D.V. Vargas, K. Sakurai, One pixel attack for fooling deep neural networks, IEEE Trans. Evol. Comput. 23 (5) (2019) 828–841.
[19] H.A. Qadir, I. Balasingham, J. Solhusvik, J. Bergsland, L. Aabakken, Y. Shin, Improving automatic polyp detection using cnn by exploiting temporal dependency in colonoscopy video, IEEE J. Biomed. Health Informatics (2019).
[20] W.-L. Chao, H. Manickavasagan, S.G. Krishna, Application of artificial intelligence in the detection and differentiation of colon polyps: a technical review for physicians, Diagnostics 9 (3) (2019) 99.
[21] D. Misra, Mish: A Self-regularized Non-Monotonic Neural Activation Function, 2019. arXiv:1908.08681.
[22] H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid, S. Savarese, Generalized intersection over union: a metric and a loss for bounding box regression, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2019) 658–666.
[23] S. Ameling, S. Wirth, D. Paulus, G. Lacey, F. Vilarino, Texture-based polyp detection in colonoscopy, in: Bildverarbeitung für die Medizin 2009, Springer, 2009, pp. 346–350.
[24] S.A. Karkanis, D.K. Iakovidis, D.E. Maroulis, D.A. Karras, M. Tzivras, Computer-aided tumor detection in endoscopic video using color wavelet features, IEEE Trans. Inform. Technol. Biomed. 7 (3) (2003) 141–152.
[25] S.Y. Park, D. Sargent, I. Spofford, K.G. Vosburgh, A. Yousif, et al., A colon video analysis framework for polyp detection, IEEE Trans. Biomed. Eng. 59 (5) (2012) 1408–1418.
[26] Q. Zhang, N. Huang, L. Yao, D. Zhang, C. Shan, J. Han, Rgb-t salient object detection via fusing multi-level cnn features, IEEE Trans. Image Process. 29 (2019) 3321–3335.
[27] D. Jha, S. Ali, N.K. Tomar, H.D. Johansen, D. Johansen, J. Rittscher, M.A. Riegler, P. Halvorsen, Real-time polyp detection, localization and segmentation in colonoscopy using deep learning, IEEE Access 9 (2021) 40496–40510.
[28] J. Li, H.-C. Wong, S.-L. Lo, Y. Xin, Multiple object detection by a deformable part-based model and an r-cnn, IEEE Signal Process. Lett. 25 (2) (2018) 288–292.
[29] S.Y. Park, D. Sargent, Colonoscopic polyp detection using convolutional neural networks, in: Medical Imaging 2016: Computer-Aided Diagnosis, Vol. 9785, International Society for Optics and Photonics, 2016, p. 978528.
[30] S. Park, M. Lee, N. Kwak, Polyp Detection in Colonoscopy Videos Using Deeply-Learned Hierarchical Features, Seoul National University, 2015.
[31] D. Bravo, J. Ruano, M. Gómez, E. Romero, Automatic polyp detection and localization during colonoscopy using convolutional neural networks, in: Medical Imaging 2020: Computer-Aided Diagnosis, vol. 11314, International Society for Optics and Photonics, 2020, p. 113143A.
[32] A. Tashk, J. Herp, E. Nadimi, Fully automatic polyp detection based on a novel u-net architecture and morphological post-process, in: 2019 International Conference on Control, Artificial Intelligence, Robotics & Optimization (ICCAIRO), IEEE, 2019, pp. 37–41.
[33] N. Tajbakhsh, S.R. Gurudu, J. Liang, Automatic polyp detection in colonoscopy videos using an ensemble of convolutional neural networks, in: 2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI), IEEE, 2015, pp. 79–83.
[34] H.-C. Shin, L. Lu, L. Kim, A. Seff, J. Yao, R.M. Summers, Interleaved text/image deep mining on a very large-scale radiology database, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2015) 1090–1099.
[35] H. Chen, D. Ni, J. Qin, S. Li, X. Yang, T. Wang, P.A. Heng, Standard plane localization in fetal ultrasound via domain transferred deep neural networks, IEEE J. Biomed. Health Informatics 19 (5) (2015) 1627–1636.
[36] H.-C. Shin, H.R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, R. M. Summers, Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning, IEEE Trans. Med. Imaging 35 (5) (2016) 1285–1298.
[37] Y. Shin, H.A. Qadir, L. Aabakken, J. Bergsland, I. Balasingham, Automatic colon polyp detection using region based deep cnn and post learning approaches, IEEE Access 6 (2018) 40950–40962.
[38] N. Tajbakhsh, J.Y. Shin, S.R. Gurudu, R.T. Hurst, C.B. Kendall, M.B. Gotway, J. Liang, Convolutional neural networks for medical image analysis: full training or fine tuning? IEEE Trans. Med. Imaging 35 (5) (2016) 1299–1312.
[39] S. Kosub, A note on the triangle inequality for the Jaccard distance, Pattern Recogn. Lett. 120 (2019) 36–38.
[40] K. Banerjee, R.R. Gupta, K. Vyas, B. Mishra, et al., Exploring Alternatives to Softmax Function, 2020. arXiv:2011.11538.
[41] M. Torres-Velázquez, W.-J. Chen, X. Li, A.B. McMillan, Application and construction of deep learning networks in medical imaging, IEEE Trans. Radiat. Plasma Med. Sci. (2020).
[42] R. Yamashita, M. Nishio, R.K.G. Do, K. Togashi, Convolutional neural networks: an overview and application in radiology, Insights Imaging 9 (4) (2018) 611–629.
[43] J. Silva, A. Histace, O. Romain, X. Dray, B. Granado, Toward embedded detection of polyps in wce images for early diagnosis of colorectal cancer, Int. J. Comput. Assist. Radiol. Surg. 9 (2) (2014) 283–293.
[44] J.Y. Lee, J. Jeong, E.M. Song, C. Ha, H.J. Lee, J.E. Koo, D.-H. Yang, N. Kim, J.-S. Byeon, Real-time detection of colon polyps during colonoscopy using deep learning: systematic validation with four independent datasets, Sci. Rep. 10 (1) (2020) 1–9.