EasyChair Preprint

№ 4421

A Dilated Convolutional Approach for Inflammatory Lesion Detection Using Multi-Scale Input Feature Fusion

Samarjeet Kaur and Nidhi Goel

October 17, 2020

A Dilated Convolutional Approach for Inflammatory Lesion Detection Using Multi-Scale Input Feature Fusion (Workshop Paper)

Samarjeet Kaur
Dept. of E & C
BVCOE
India
sammarjeetkaur@gmail.com

Nidhi Goel, Senior Member, IEEE
Dept. of E & C
IGDTUW
India
nidhi.iitr1@gmail.com

***Abstract*—The present manuscript proposes a novel CNN architecture to detect inflammatory lesion abnormality in Wireless Capsule Endoscopy (WCE) images. Such images encompasses a wide range of lesions and hence early diagnosis can be of vital importance. The proposed model learns the collective features of various inflammatory lesion subgroups and aggregates that information to solve a binary classification problem by distinguishing between normal and abnormal frames. The proposed model has one primary and three secondary branches. The primary branch resembles a generic CNN model with convolution and max-pooling layers whereas the secondary branches consist of dilated convolution layers and max-pooling layers. The proposed model fuses the multi-scale input context at varying dilation rates with different levels of the primary branch. This enhances feature quality by merging dominant global features with the local input context at multiple scales without any loss of resolution. The performance of the proposed model has been assessed using various objective evaluation metrics. The preliminary experiments indicates that the proposed model outperforms state-of-the-art models and exhibits an accuracy of 97.9%, sensitivity of 96%, specificity of 99%, ROC-AUC of 1 and Precision recall AUC of 99.7%.**

***Keywords*-Convolution neural network; wireless capsule endoscopy; lesion detection;**

Wireless Capsule Endoscopy (WCE) is a non-invasive alternative to traditional endoscopic procedures [1]. It is used for detecting abnormalities in the small bowel. The procedure begins with the ingestion of WCE capsule. The capsule is then propelled through the gastrointestinal tract by peristalsis [2]. The device can record up to $8-12$ hours of video sequences during the procedure. One WCE procedure can produce more than $60,000$ frames and its analysis requires around $120-180$ minutes of intense attention and focus by a clinician [1], [2].

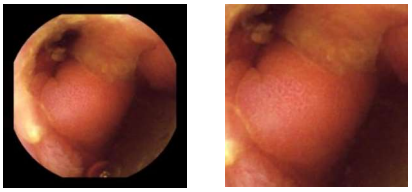
Lesion is one of the most frequently found occurring abnormality in the small bowel and often diagnosed as the source of bleeding [3]. Therefore, early diagnosis can prove to be a very crucial first step towards faster and safer recovery. Recently, a lot of computer aided methods have been developed to reduce WCE screening time and lessen the burden on clinicians by developing algorithms for automating the anomaly detection process [4], [5], [6], [7]. Deep learning architectures like Convolutional Neural Networks have shown promising results in detecting abnormalities in various kinds of endoscopy images [8], [9].

Single Shot Multibox Detector (SSD), which is a deep neural network architecture, has been used for detection of ulcers and erosions in WCE images [3]. Similarly, SSD algorithm was also applied for detection and classification of protruding lesions including polyps, SMTs, epithelial tumours, nodules, and venous structures [10].

Lack of availability of labelled medical data poses a huge problem for designing efficient, robust and well-generalizable deep learning models for abnormality detection. To overcome this limitation, the possibility of fine tuning of pre-trained models like GoogLeNet, AlexNet etc. has also been explored [11]. Annotations pose a limitation but often image level labels are feasible and available whereas graphic annotations are extremely time consuming and hard to obtain. The use of image level annotations with weakly supervised CNNs has been explored for automating the diagnosis for inflammatory lesion detection [12].

Similarly, a combination of CNN and Long Short Term Memory (LSTM) has been explored for leveraging temporal relationships between the video sequenced frames for detection of arteriovenous malformations (AVM) [13]. Literature has pointed out that non-RGB colour spaces extract more colour and texture features as compared to RGB colour space [14]. This led to the use of CIELab and RGB input colour spaces in conjunction with transfer learning approach which had GoogLeNet as the underlying architecture for lesion detection in WCE images [15].

Thorough literature review revealed that CNN based methods have been extensively explored for tackling abnormality detection in WCE images [5], [15]. Designing CNN based models right from the scratch comes with an advantage of architectural flexibility by making a trade-off between model depth and its performance. Empirical data shows that performance improves with increasing depth of the model. Though the performance improves with depth; deep models tend to over-fit very quickly if the dataset is not large enough. Moreover, for a robust model it is



(a) (b)

Figure 1. (a) Original image (b) Cropped image

important that the data supplied during training encompasses enough variations across samples so that the model learns all possible patterns/variations that may occur.

The present paper, thus, proposes a novel CNN based architecture which is deep enough to learn the right features and compact enough for making real-time deployment feasible. The proposed model performs binary classification in order to distinguish between normal and abnormal frames. The abnormal class contains lesions belonging to various subcategories, *viz.*, inflammatory lesions, including ulcers, mucosal aphthae, mucosal erythema, mucosal cobblestone and luminal stenosis.

In the proposed work, input images are first converted to LUV colour space and the L channel is histogram equalized. Then the image is simultaneously fed into four branches (connected to a common input). Out of the four branches, one is primary and others are secondary. The problem of limited data has been solved by using various augmentation techniques. The proposed model also combats the problem of resolution loss by preserving the input context using dilated convolution which makes the model capable of identifying both large and small anomaly regions with the same effectiveness. In order to further enhance the feature quality, we extract multi-scale features via dilated convolution from the input context.

The dataset was divided into train, test and validation sets. The model is objectively evaluated using test set for various quantitative metrics. The obtained accuracy, sensitivity, specificity and ROC-AUC scores are 97.9%, 96% 99% and 1, respectively.

The rest of this paper is organized as follows. Section II discusses the preparation of dataset while Section III presents the proposed CNN based framework, model training and parameter setting. Section IV discusses the experimental analysis and the paper has been concluded in Section V.

I. Dataset Preparation

The present section discusses the preparation of dataset which primarily includes obtaining the dataset, the pre-processing operation and augmentation applied on the dataset.

A. Dataset

All the experiments in the proposed manuscript were conducted on real clinical data generated by 252 WCE procedures performed at Royal Infirmary of Edinburgh (University Hospital and referral centre for WCE for the Southeast of Scotland, UK) [16]. A total of 1372 images were obtained from the video sequences and classified by a research group as follows: a) vascular lesions, that includes angiectasia and/or intraluminal bleeding; b) inflammatory lesions, including ulcers, mucosal aphthae, mucosal erythema, mucosal cobblestone and luminal stenosis; c) lymphangiectasia, including nodular lymphangiectasis, chylous cysts, punctuate lymphangiectasis, and d) polypoid lesions [16]. For the proposed experiments, only a subset of this dataset belonging to the target class (lesions) was used. Data pertaining to target class has 204 inflammatory lesion frames and 57 normal frames, that were used for abnormal and normal classes, respectively.

B. Pre-processing

The circular aperture of a WCE camera lets the important information resides in the central region of a WCE image while the black borders around this central region indicates the date and time stamp or the capsule specifications. Analysis indicated that this border area increases the size of the image without adding any meaningful information. Thus, the proposed work carefully implements cropping as a part of pre-processing step to reduce redundancy without losing on the important medical information. Cropping reduces the original size of image to 240×240. The original image along with its cropped image has been indicated in Figure 1.

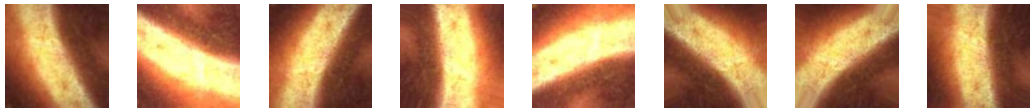
C. Data Augmentation

The resulting subset containing abnormal and normal frames of lesion amounted to only 261 images in total. However, this 261 image dataset is insufficient for training of a Deep Learning Model, which requires a large dataset with a lot of variance within it. Thus, the chosen dataset is augmented using different methods to make the dataset suitable for training of the proposed deep learning model. The data augmentation step is done primarily for better training of the proposed model, that consequently leads to a generalized model that prevents overfitting.

For data augmentation, the following transformations have been applied: rotation shift, brightness shift, horizontal and vertical flip. After augmentation, equal number of frames were generated for both the classes *i.e.* 4000 frames for each class. This makes the complete dataset of 8000 images. An original image and its respective augmented images have been indicated in Figure 2.

II. Proposed Framework

The present section discusses the motivation behind the proposed CNN based architecture, discusses the model



(a) (b) (c) (d) (e) (f) (g) (h)

Figure 2. (a) Original image (b-h) Augmented images

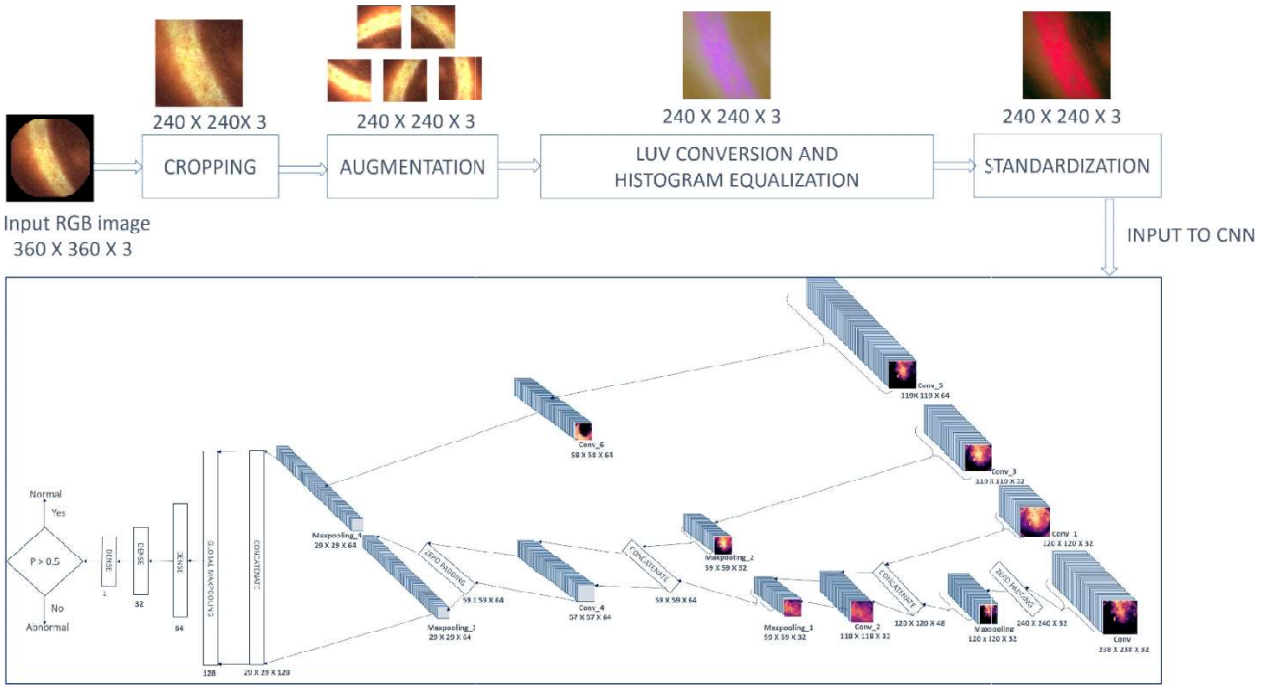


Figure 3. Block diagram of proposed CNN based architeture

parameters and the parameter settings for the proposed architecture. The block diagram of the complete framework is indicated in Figure 3.

A. *Overview of the Proposed Model*

While designing CNN architecture for predictions on medical data, the size of the model and the information retention capacity are the critical parameters and are proportional to each other. Decreasing the former evidently decreases the latter. For an efficient deep learning network, it is essential to have a large network that is trained on a huge database. A large sized network increases the information retention capacity of the network. This can be considered as an advantage in some cases, but it comes at a cost of increased computational time, which cannot be compromised during real time scenarios. To have an edge over the computational time, a trade-off is required that reduces the size of the network without losing on the information retention capacity.

Literature has revealed that conventional CNNs perform down-sampling operations and dimensionality reduction using pooling layers (or subsampling layers) in order to capture larger context areas [17]. Pooling layers help expand the receptive field at the cost of resolution [18]. Pooling layers reduce the number of parameters by retaining only a small amount of information which is guided by the pooling algorithm being used. This leads to loss of resolution which is not favourable in the context of medical images because it may completely change the interpretation of the diseased region.

On the contrary, dilated convolution preserves spatial information better than pooling layers, thereby improving the quality of generated feature maps [19]. Also, dilated convolution does not compromise upon the image resolution

while reducing the network parameters. Thus, the proposed model has explored the combination of dominant feature extraction ability of max-pool operation and resolution preserving characteristic of dilated convolution to improve the anomaly detection and classification capability of the network.

B. Model training and Parameter setting

The model pipeline has two parts: on-the-fly pre-processing and the proposed CNN based model. On-the-fly pre-processing is performed by a pre-processing function. Firstly, images are loaded into the memory in small batches given the memory constraints. Every image is processed by the pre-processing function before it is fed into the model. The function first brings the image pixel values within [0,1] range and converts the input RGB colour space into LUV followed by histogram equalization on the L channel.

In the colour space transformation step, LUV was chosen as the input colour space as it separates chromaticity from luminance thereby making it invariant to lighting conditions. Moreover, compared to RGB, LUV is more close to human perception of lightness and chromaticity. All these factors make LUV a suitable choice for lesion detection. The luminance channel of the LUV image was histogram equalized in order to enhance global image contrast by making the intensity distribution uniform over the image. After histogram equalization, the obtained image undergoes standardization *i.e.* mean centering and normalization by standard deviation.

The proposed model was implemented using Keras with Tensorflow backend. The learning rate was set to 0.0001 with a batch size of 10. ReLu activation was used for all the convolution layers and sigmoid activation was used in the last three fully connected layers. The model's hyper-parameter tuning was performed using the trial and error approach. Model was trained for 100 epochs over a duration of 7 hours. The experiments were performed on a system with these specifications: Intel Core i5-8300H processor with 24GB system RAM, NVIDIA GeForce GTX 1050Ti with 4GB RAM graphics card.

C. Proposed Model Architecture

The proposed model is a single input multi-branch CNN having one primary branch and three secondary branches. The primary branch has three regular convolution layers with a fixed kernel size of 3×3. Each Conv layer is followed by a 2×2 max-pool layer, with the exception of two intermediate zero-padding layers of $(1, 1)$ for ensuring a symmetric output. Max-pool layers are followed by concatenation layers which concatenate the output of the secondary branches at different levels of the main branch. All secondary branches are connected to the main input and each of them comprises one or multiple dilated convolution layers with kernel size ranging from 2×2 to 4×4 and a fixed dilation rate of 2. ReLu activation has been used as the activation function for every convolution layer.

It is to be noted that each of the secondary branch's output is being concatenated with the primary branch at different levels, consequently successive secondary branches needed multiple down-sampling layers. The first secondary branch has one dilated Conv layer. The second secondary branch consists of one dilated Conv and one 2×2 max-pool layer. The third secondary branch has two dilated convolution layers and one 2×2 max-pooling layer. For this branch, multiple dilated convolution operations were chosen over multiple max-pool operations by taking into consideration the loss of information that is accompanied by the latter. Two stacked dilated convolution layers ensured preservation of resolution to a great extent even after significant down-sampling.

The secondary branches helps to preserve the original input context and extract features at multiple scales by using different sized kernels. The extracted features from the secondary branch are repeatedly concatenated with input's dilated convolution feature maps of the primary branch at deeper levels. This ensures that both the coarse and fine details are taken into account.

The last concatenation layer is followed by a global max-pooling layer which acts as the flattening layer in this case. It only retains the maximum feature per feature map thereby producing a 1D vector. Three dense layers are stacked one after the other in the end with the last layer having a single node with sigmoid activation function to give a class probability as output.

III. Experimental Analysis

This section discusses the detailed experimental analysis on the prepared dataset of 4000 abnormal and 4000 normal frames. This data was divided into training, validation and test set. The training set comprised of 6000 frames in total (3000 normal and 3000 abnormal), while the validation and test set had 1000 frames each (500 frames per class).

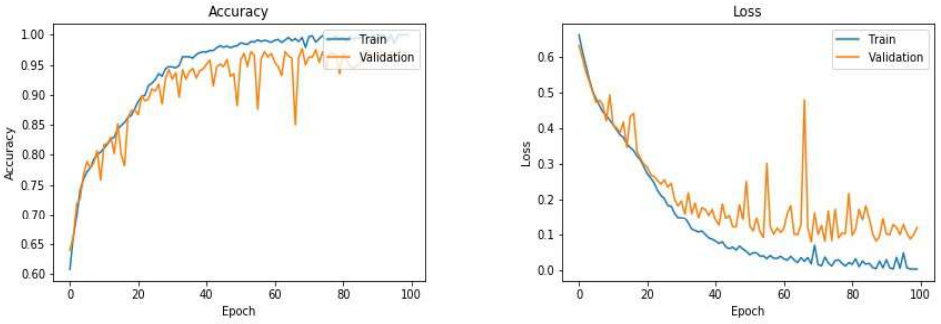
A. Experiment Methodology

The proposed model was trained on the training dataset of 6000 images and the experimental analysis has been performed on the test set. The results indicated in the tables refers to the results obtained on the test dataset. The pre-processing function takes cropped images as input (pre-processing), normalizes and performs colour space conversion from RGB to LUV followed by histogram equalization of the Luminance or 'L' channel (on-the-fly-pre-processing), this input is then fed into the model and a prediction in terms of class probability is obtained.

The sigmoid activation function used in the last layer of the model predicts the probability of a frame being normal. Therefore, if the output probability is 'p' then the probability

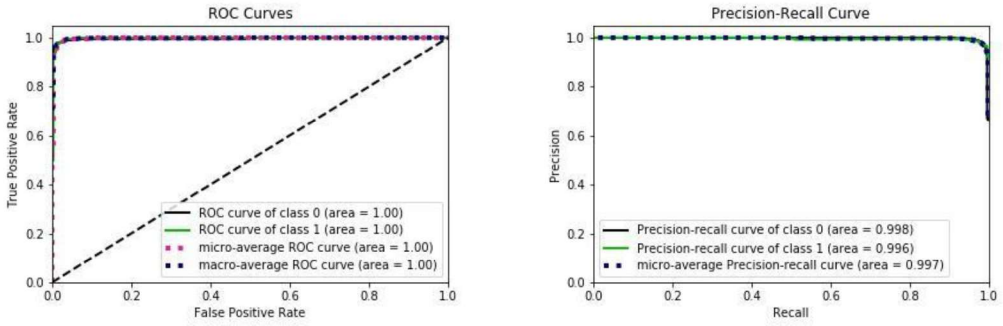
Table I
EVALUATION RESULTS

Class	Precision	Recall	FI Score	Sensitivity	Specificity	ROC AUC	Precision-Recall AUC
Abnormal	0.99	0.96	0.98	0.96	0.99	1	99.8
Normal	0.97	0.99	0.98	0.99	0.96	1	99.6



(a) (b)

Figure 4. Accuracy and Loss plot for the proposed CNN based architeture



(a) (b)

Figure 5. ROC and Precision Recall Curves for the proposed CNN based architeture

of normalcy is 'p' and probability of the given frame being abnormal is '1-p'.

The output probability is converted into a binary label by using a threshold value of 0.5. This threshold value is chosen based on empirical observations. Hence, if the output probability is greater than 0.5 then the frame is classified as normal else it is classified as abnormal.

B. Evaluation Results

The performance scores on the test dataset have been indicated in Table I. Classification accuracy of 97.9% has been obtained on the test set. Precision of 0.99 and 0.97 has been obtained for abnormal and normal frames. Similarly, recall of 0.96 and 0.99 has been obtained for abnormal and normal frames. An F1 score of 0.98 is obtained for both abnormal and normal frames. The achieved accuracy, sensitivity, specificity, ROC-AUC and precision recall-AUC is also indicated in the Table I, which indicates the efficacy of the proposed framework.

The model was trained for 100 epochs, and the obtained accuracy and loss for training/validation have been indicated in Figure 4. In both the plots, training and validation curves are closely following each other, indicating that the model is neither overfitting nor under-fitting. There are some spikes in the validation loss curve, which may be due to the mini-batch gradient descent.

Area under curve (AUC) for ROC and Precision-Recall curves have been shown in Figure 5 which indicates the

Table II
COMPARATIVE ANALYSIS

-	Saito *et al.* [10]	Aoki *et al.* [3]	Alaskar *et al.*[11]		Spiros *et al.* [12]	Proposed
Underlying algorithm/Model	Single Shot Multibox detector	Single Shot Multibox detector	GoogleLeNet	Alexnet	Weakly supervised CNN	Dilated Convolution Approach
Dataset source	Sendai Kousei Hospital, The University of Tokyo, and Hiroshima University Hospital	The University of Tokyo Hospital, Japan	Dr. Khoroo's Medical Clinic/Trust		KID Dataset	Royal Infirmary of Edinburgh (University Hospital and referral centre for WCE for the southeast of Scotland, UK) [16]
Dataset details	Training-30,584 frames, Testing-17,507 frames	Training-5360 frames, Testing-10440 frames	Training-336 frames, Testing- 105 frames)	Training-336 frames, Testing- 105 frames	Training-400 frames, Testing- 54 frames	Training-6000 frames, Validation-1000 frames, Testing-1000 frames
Lesion subcategories	polyps, epithelial tumors, submucosal tumors (SMT), nodules, and venous structures	Erosions and Ulcers	Ulcers	Ulcers	-	inflammatory lesions, including mucosal aphthae, mucosal erythema, mucosal erythema, ulcers, mucosal cobblestone, and luminal stenosis
Augmentation	Performed	Not performed	Not performed	Not performed	Not performed	Performed
Accuracy(%)	-	90.8	100	100	90.2	97.9
Sensitivity (%)	90.7	88.2	100	100	92.6	96
Specificity (%)	79.8	90.9	100	100	88.9	99
AUC-ROC	0.911	0.958	1	1	-	1
Testing (sec/frame)	0.00303	0.022	-	-	-	0.0175
Trainable parameters	-	-	-	-	-	131,889

degree of differentiability between the classes present in the data. The plotted ROC curve indicates that for True positive Rate of 1, there exists a point on the curve where False positive rate is 0, indicating an ideal performance. Similarly, for Precision-Recall curve, ideal performance is indicated where both precision and recall values are 1. Greater the AUC, better is the model performance. A perfect model has an AUC of 1.

From the obtained graphs, it can be clearly seen that for both ROC and Precision-Recall curves, the model is achieving near perfect AUC score. For ROC, both normal (label 1) and abnormal (label 0) class achieve an AUC of 1 whereas in Precision –Recall curve abnormal class shows better performance with an AUC of 0.998 as compared to the normal class with AUC of 0.996.

The comparative analysis of the proposed model has been done with state-of-the-art works. For the comparative analysis, existing work on specific anomaly class i.e. lesions have been considered. Only the best case performance scores obtained on the test/validation data (as stated in the respective literature), are reported in the Table II.

The architecture proposed by Aoki *et al.* [3] and Saito *et al.* [10] uses Single Shot Multibox detector (SSD) as its underlying algorithm and required region level annotations for binary classification. The proposed model however requires image level annotations. Though, a weakly supervised CNN

has been proposed using image level annotations [12]; the parametric values are less than the values obtained for the proposed method (Table II).

Two widely used pretrained models namely GoogLeNet and AlexNet were fine-tuned on a dataset of 256 abnormal and 80 normal ulcer frames [11]. The reported accuracy, sensitivity and specificity scores were 100%, 100% and 100% with an AUC of 1. However, it is to be noted that the validation was performed on a very small subset of the data consisting of only 80 abnormal and 25 normal images unlike the proposed model which was validated on 1000 augmented images which captured a very wide range of variations. In order to obtain representative scores, the test set should encompass significant proportion of the possible variations that might occur in real time data. Moreover, the architectures GoogLeNet and AlexNet have parameters roughly of the order 4×10^6 and 60×10^6 [20], which is ten times more than the total parameters of the proposed model. Using a pre-trained architecture also eliminates the possibility of altering the architecture in the future to suit a specific use case.

A thorough comparative analysis presented in Table II depicts that the proposed method outperforms existing state-of-the-art methods in terms of sensitivity, specificity, AUC and per frame testing times.

IV. Conclusion

Recent years have seen the exponentially growing use of deep learning methods for automating the detection of abnormalities in a various medical images. The proposed framework is a lightweight CNN model for inflammatory lesion detection that uses dilated convolution for input context preservation with significantly fewer parameters. Multiscale features are extracted from the input image at different dilation rates and they are fused with the main network at different depths.

The proposed CNN based model requires only image level annotations for training, thus, eliminating the need for time consuming graphical annotations for binary classification of lesion detection The promising performance of the proposed method paves way for further study of its performance on a larger dataset with multiple disease classes.

Acknowledgment

The authors would like to express their sincere gratitude towards Dr. Deepak Gunjan, Department of Gastroenterology and Human Nutrition Unit, All India Institute of Medical Sciences, New Delhi for all the insight about the endoscopic images and the various abnormalities. The authors also acknowledges the SMDP-C2SD project, Ministry of Electronics and Information Technology, Government of India for the research grant.

References

[1] Enns, Robert A. *et al.* "Clinical practice guidelines for the use of video capsule endoscopy," Gastroenterology, vol. 152, issue 3, pp. 497-514, 2017.

[2] Gavriel Iddan, Gavriel Meron, Arkady Glukhovsky, and Paul Swain, "Wireless capsule endoscopy," Nature, vol. 405, 2000.

[3] Aoki, Tomonori, *et al.* "Automatic detection of erosions and ulcerations in wireless capsule endoscopy images based on a deep convolutional neural network," Gastrointestinal endoscopy, vol. 89, issue 2, pp. 357-363, 2019.

[4] Y. Fu, W. Zhang, M. Mandal, and M. Q.-H. Meng, "Computer-aided bleeding detection in WCE video," IEEE J. Biomed. Health Informat., vol. 18, no. 2, pp. 636–642, Mar. 2014.

[5] Jun-Yan He, Xiao Wu, Yu-Gang Jiang, Qiang Peng, and Ramesh Jain, "Hookworm detection in wireless capsule endoscopy images with deep learning," IEEE Transactions on Image Processing, vol. 27, no. 5, pp. 2379-2392, May 2018.

[6] S. Charfi, M. El Ansari and I. Balasingham, "Computer-aided diagnosis system for ulcer detection in wireless capsule endoscopy images," IET Image Processing, vol. 13, no. 6, pp. 1023-1030, 5 2019.

[7] D.K. Iakovidis, A. Koulaouzidis, "Automatic lesion detection in wireless capsule endoscopy – A simple solution for a complex problem," Proc. IEEE International Conference on Image Processing (ICIP), Paris, France, 2014, pp. 2236-2240.

[8] D. K. Iakovidis, S. V. Georgakopoulos, M. Vasilakakis, A. Koulaouzidis and V. P. Plagianakos, "Detecting and locating gastrointestinal anomalies Using deep learning and iterative cluster unification," IEEE Transactions on Medical Imaging, vol. 37, no. 10, pp. 2196-2210, Oct. 2018.

[9] Soffer, Shelly, *et al.* "Deep learning for wireless capsule endoscopy: a systematic review and meta-analysis," Gastrointestinal Endoscopy, 2020 (in Press).

[10] Saito, Hiroaki, et al. "Automatic detection and classification of protruding lesions in wireless capsule endoscopy images based on a deep convolutional neural network," Gastrointestinal Endoscopy, vol. 92, issue 1, pp. 144-151, July 2020.

[11] Haya Alaskar, Abir Hussain, Nourah Al-Aseem, Panos Liatsis, and Dhiya Al-Jumeily,"Application of convolutional neural networks for automated ulcer detection in wireless capsule endoscopy images," Sensors, vol. 19, issue 6, 2019.

[12] Spiros V. Georgakopoulos, Dimitris K. Iakovidis, Michael Vasilakakis, Vassilis P. Plagianakos, and Anastasios Koulaouzidis, "Weakly-supervised convolutional learning for detection of inflammatory gastrointestinal lesions," IEEE International Conference on Imaging Systems and Techniques, October 2016.

[13] S.S. Rajagopalan, S.C. Rice, P.R. Slawinski, P. Valdastri, and K.L. Obstein, "Evaluation of an automated lesion detection platform for wireless capsule endoscopy: a novel approach utilizing video-based machine learning temporal relationships," Gastrointestinal Endoscopy, vol. 89, issue 6, 2019.

[14] Eric Dubois, “The structure and properties of color spaces and the representation of color images,” Morgan & Claypool, 2009.

[15] Yejin Jeon, Eunbyul Cho, Sehwa Moon, Seung-Hoon Chae, Hae Young Jo, Tae Oh Kim, Chang Mo Moon, and Jang-Hwan Choi, “Deep convolutional neural network-based automated lesion detection in wireless capsule endoscopy,” International Forum on Medical Imaging in Asia 2019. Vol. 11050.

[16] Dimitris K Iakovidis and Anastasios Koulaouzidis, “Automatic lesion detection in capsule endoscopy based on color saliency: Closer to an essential adjunct for reviewing software,” Gastrointest Endoscopy, vol. 80, issue 5, pp. 877-83, 2014.

[17] Rikiya Yamashita,Mizuho Nishio, Richard Kinh Gian Do & Kaori Togashi, “Convolutional neural networks: an overview and application in radiology,” Insights into Imaging, vol. 9, pp. 611 - 629, 2018.

[18] Hamaguchi, Ryuhei and Fujita, Aito and Nemoto, Keisuke and Imaizumi, Tomoyuki and Hikosaka, Shuhe, “Effective use of dilated convolutions for segmenting small object instances in remote sensing imagery,” vol. 9, 2017.

[19] Yuqian Zhang, Guohui Li,Jun Lei and Jiayu He, ”FDCNet: Frontend-backend fusion dilated network through channel-attention mechanism,” Applied Science, vol. 9, issue 17, 3466, 2019.

[20] Sudha, K. K., and P. Sujatha, “A qualitative analysis of Googlenet and Alexnet for fabric defect detection.” Int. J. Recent Technol. Eng. 8 (2019): 86-92.