A study on endoscopy artefact detection based on deep learning

Han Zhang
University of Electronic Science and Technology of China
Prince Adjei
University of Electronic Science and Technology of China
Xiaoming Hu
University of Electronic Science and Technology of China
Xiaotong Wang
University of Electronic Science and Technology of China
Yeting Hu
University of Electronic Science and Technology of China
Tao Gan
Digestive Endoscopic Center of West China Hospital, Sichuan University
Nini Rao (raonn@uestc.edu.cn)
Center for Informational Biology, University of Electronic Science and Technology of China

Research Article

Keywords: endoscopic image, artefact detection, artificial intelligence, feature pyramid network

Posted Date: September 14th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1995620/v1

A study on endoscopy artefact detection based on deep learning

Han Zhang[1,2], Prince Adjei[1,2], Xiaoming Hu[1,2], Xiaotong Wang[1,2], Yeting Hu[1,2], Tao Gan[4] and Nini Rao[2,3*]

[1]Key Laboratory for Neuro-information of Ministry of Education, University of Electronic Science and Technology of China, Chengdu, 610054, Sichuan, China.

[2]School of Life Science and Technology, University of Electronic Science and Technology of China, Chengdu, 610054, Sichuan, China.

[3*]Center for Informational Biology, University of Electronic Science and Technology of China, Chengdu, 610054, Sichuan, China.

[4]Digestive Endoscopic Center of West China Hospital, Sichuan University, Chengdu, 610017, Sichuan, China.

*Corresponding author(s):Nini Rao, Ph.D., Center for Informational Biology, University of Electronic Science and Technology of China, Chengdu 610054, China and School of Life Science and Technology, University of Electronic Science and Technology of China, Chengdu 610054, China. Tel: +86-028-83208238; Fax: +86-028-83208238; E-mail(s): raonn@uestc.edu.cn;

Contributing authors: zhanghan_ella@163.com; prince.e.adjei@gmail.com; 1344584494@qq.com; 1079480359@qq.com; 2465297030@qq.com; gantao111@aliyun.com;

Abstract

Endoscopic images are generally corrupted by various artefacts, such as blur caused by the irregular motion of endoscopists or organ

peristalsis, preventing endoscopists from performing efficient and high-accuracy diagnosis and limiting the utility of computer-aided diagnosis in endoscopy. As a result, an endoscopy artefact detection algorithm that is both robust and reliable is critical in endoscopic imaging. This study proposes an endoscopy artefact detection method based on an improved Faster Region-based Convolutional Neural Network (Faster R-CNN) detector, simply named Faster RFD3-CNN. We introduce deformable convolutions and a feature pyramid network (FPN) to strengthen the model's adaptability and feature extraction ability. Furthermore, class-aware non-maximum suppression (NMS) and false positive elimination are utilized to filter out some bounding boxes with low confidence or overlap. The experimental results on the public access EndoCV2020 data set show that the mean average precision (mAP) can reach 42.66% and the intersection of union (IoU) is 31.61%. The mAP is increased by nearly 15% compared with the existing methods. Compared with the basic Faster R-CNN, mAP and IoU are improved by 7.39% and 11.19%, respectively. To sum up, the proposed method in this paper has superior performance in endoscopy artefact detection.

Keywords: endoscopic image, artefact detection, artificial intelligence, feature pyramid network

Introduction

Digestive endoscopes in clinical use are generally made of metal or plastic tubes of different thicknesses, lengths and shapes, with a light source, camera and other systems at the tip. Endoscopists can examine the internal conditions of the digestive tract from the external display connected to the endoscopes. Endoscopy is widely used in the early screening, treatment, and minimally invasive surgery of digestive tract diseases and is considered the gold standard for detecting digestive tract abnormalities and cancer. However, endoscopic images differ from natural images; the lack of optimal reflected light, inevitable organ peristalsis, and profound differences in organ shape and surface texture of different digestive tract segments and body fluids are sources of artefacts [1].One study showed that more than 60% of endoscopic video frames and nearly 70% of endoscopic video sequences are damaged by multiple artefacts, including floating fluid, specularity, bubbles, chyme, and other artefacts [2].Typically, such artefacts mask or mimic pathologies and lesions, potentially leading to increased missed detections and misdiagnosed patients.

Endoscopy artefacts can roughly be divided into eight classes; overexposure and underexposure of image areas caused by changes in lighting and organ topology (referred to as "saturation" and "low contrast" respectively), blur caused by irregular hand movements (of endoscopists) and local organ movement, specularity caused by light reflected from smooth organ surfaces, and

shielding with instruments for minimally invasive surgery and biopsy, internal cavity bleeding, gas forming dense bubbles in the digestive tract distort the appearance of tissues, and other general artefacts.Zhang et al. [3] showed that the performance of polyp tracking detectors in colonoscopy would be affected by endoscopy artefacts due to the similarity between some types of artefacts and lesions, and the inherent limitations of object detectors based on artificial intelligence. Further, another study reported an increase in the sensitivity of the lesion detector with artefacts and a corresponding increase in the false positive rate, affecting the model's overall accuracy [4].Artefacts affect automated disease detection, reporting and surgical planning. Detecting and locating endoscopy artefacts is thus a fundamental challenge in endoscopic imaging.

Current endoscopy artefact detection strategies can roughly be divided into handcrafted feature-based and deep learning-based methods. Handcrafted features refer to properties derived using predefined algorithms based on expert knowledge using the information present in the image itself [5].LBP [6, 7],SIFT [8], and HOG [9] are typical and universally known algorithms for extracting handcrafted features.Queiroz et al. [10] used a nonlinear enhancement filter to highlight the differences between the reflection area and background of endoscopy images. Segui et al. [11] combined color histograms and texture features in the small intestine to segment bubbles. Zhao et al. [12] proposed a general framework of wireless capsule endoscopy based on the Hidden Markov model, combined with MPEG7 features to detect bubbles, bile, feces and food residues in the digestive tract. The endoscopy artefact detection method based on handcrafted features focuses on the research of specific artefacts, and a single imaging modality on specific organs, making dealing with all artefacts challenging.

On the other hand, deep learning methods can mine deeper and more abstract features than handcrafted methods. Yang et al. [13] improved a Cascade R-CNN [14] by adding a FPN module, successfully detecting specularity, saturation, artefact, blur, contrast, bubbles and instrument. Polat et al. [15] combined Faster R-CNN [16], Cascade R-CNN and RetinaNet [17] with a consensus integration strategy to detect eight artefacts (specularity, saturation, artefact, blur, contrast, bubbles, instrument and blood). This method won first place in the official test set of EAD2020 with a performance of 20.31% for mAP and 32.85% for IoU. However, there is a problem of slow detection speed caused by the complexity of framework, which hardly caters to the needs of real-time detection. From the perspective of data imbalance, Artunc et al. [18] trained three independent YOLOv5 with the raw data, minority augmented data, and majority augmented data, respectively, for detecting endoscopy artefacts.

To improve the accuracy of artefact detection, this work takes Faster R-CNN as the core framework. Inspired by the idea that deformable convolution introduces a 2D offset at each square sample point, allowing the sampling network to deform freely and effectively find artefacts [19], we incorporate the feature pyramid and deformable convolution into the standard Faster R-CNN,

constructing a novel Faster RFD3-CNN. Finally, the novel model has been altered to use class-aware NMS and a false positive elimination for the post processing.

Materials and Methods

Materials

The EAD2020 dataset [1, 2, 20, 21], containing 2200 single frames, 232 sequence frames and 99 frames for the out-of-sample generalization task was used. Table 1 summarizes and shows two obvious challenges with datasets: (1) An imbalance in the distribution of various artefacts is unavoidable. (2) Multiple and dense artefacts may coexist in a single frame.

Table 1 Statistics of the dataset used in this work(%)

Class	Annotations Proportion	Image Proportion
Blur	2.52	25.21
Bubbles	18.12	27.80
Specularity	36.42	44.04
Saturation	4.54	44.16
Contrast	6.20	66.12
Instrument	1.99	16.16
Blood	1.65	8.35
Artefact	28.54	78.58

The train and test splits are 80% and 20% of the overall data set with the strategy of stratified sampling. All frames are uniformly resized to 512×512, similar to [15]. The data augmentation techniques utilized for training include horizontal flipping, 90°, 180°, 270° rotations, random zoom out and color jittering.

Methods

Detection model design

The proposed model is an improvement to the Faster R-CNN, named Faster RFD3-CNN, shown in Fig. 1. To enhance the detection of small region targets, the last three pieces of the original ResNet50 structure are replaced with deformable convolution (denoted D) to automatically calculate the offset of each point and take features from the most suitable place for convolution. In addition, we utilized a feature pyramid structure, shown in Fig. 2, which involves a bottom-up pathway, a top-down pathway and lateral connections to integrate multi-scale feature maps.

The loss function consists of two components: regression losses and classification losses. An essential difference between this network and Faster R-CNN is the use of focal loss in multiple final classifications. The focal loss is an

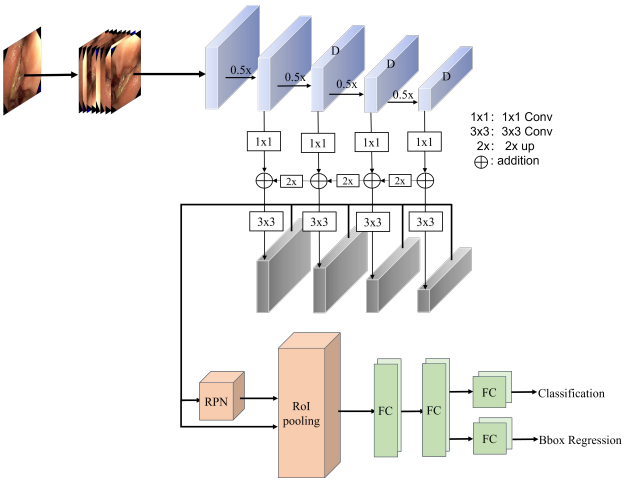

Fig. 1 The framework of Faster RFD3-CNN

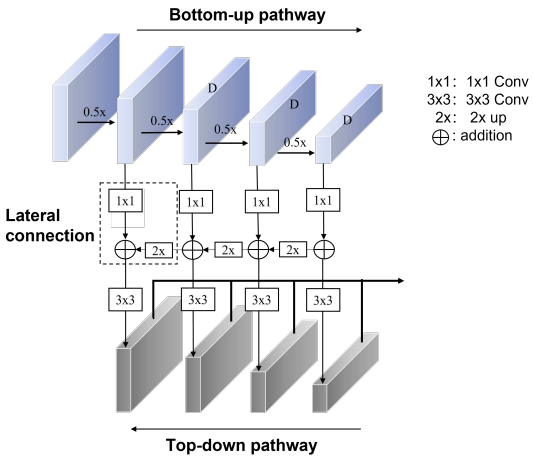

Fig. 2 A building block illustrating the bottom-up pathway, lateral connection and top-down pathway

extension to cross-entropy loss, which increases the classification accuracy of CNN by adding a weighting factor $\alpha \in [0,1]$ to balance the importance of positive/negative examples and by introducing a modulating factor $(1-p)^{\gamma}$ with tunable focusing parameter $\gamma \geq 0$ to reduce the loss contribution from easy examples, that is:

$$L_f(p,y) = \begin{cases} -\alpha(1-p)^{\gamma} log p & y=1 \\ -(1-\alpha)p^{\gamma} log(1-p) & y=0 \end{cases} \quad (1)$$

In equation 1 above, y specifies the ground truth, and p is the prediction output.

Post processing

Here, non-maximum suppression (NMS) is performed independently on each artefact in Table 1, referred to as class-aware NMS. Although class-aware NMS discards partially overlapping bounding boxes, it can still not eliminate false positive bounding boxes with a low confidence level. Some solutions are proposed in order to remove false positive bounding boxes. Because of their varying natural frequencies in endoscopic images, different types of artefacts have different prediction confidence levels. This study provides a false positive elimination method for determining a class-specific confidence threshold. Inspired by the widely used 3σ principle [22] and other false positive elimination methods [15], we calculate the thresholds using the equation (2) and (3):

$$\sigma_{cls} = \sqrt{\frac{1}{N_{cls}} \sum_{i=1}^{N_{cls}} \left(confidence_{cls,i} - \mu_{cls}\right)^2} \tag{2}$$

$$\mu_{cls} - 3\sigma_{cls} \leq threshold_{cls} \leq \mu_{cls} + 3\sigma_{cls} \tag{3}$$

Where N_{cls} is the amount of prediction bounding boxes of a certain kind of artefact, $confidence_{(cls,i)}$ is the prediction confidence of prediction bounding box, μ_{cls} is the average prediction confidence in a certain kind of artefact. The prediction bounding box will be removed if its confidence is not within the range of threshold.

Implementation details

In order to speed up training, we use the pre-trained weights on ImageNet to initial backbone ResNet50. The stochastic gradient descent is chosen as the optimizer with a momentum of 0.9 and weight decay of $10^{-}4$. To keep the distribution stable and the model stable when the model converges, the warm-up strategy is adopted to increase the learning linearly to 0.02 in the first 1000 iterations and then change according to the cosine function, whose period is set as 10 epochs. 30 epochs are trained, and the batch size is 8. As for anchors, we use five anchor scales of 32^2,64^2,128^2,256^2,512^2 and seven different anchor ratios: 0.1, 0.25, 0.5, 1.0, 2.0, 4.0, and 10.0.

Results

All experiments were conducted in Google Colab, and the programming language implemented is Python3.7. The deep learning framework adopted is PyTorch1.10.0 (https://pytorch.org/)

Evaluation metrics

Here we use well-defined metrics by the community of object detection for evaluation. The Intersection of Union (IoU), mean average precision (mAP) and a weighted score of mAP and IoU. Furthermore, COCO metrics are included.

1. IoU is obtained by the ratio of the intersection and union of the ground truth (denoted A) and the predicted bounding boxes (denoted B).

$$IoU = \frac{A \cap B}{A \cup B} \tag{4}$$

2. In terms of numbers of true positives (TP), false positives (FP) and false negatives (FN), precision (p) is defined as $\frac{TP}{TP+FP}$ and recall (r) is defined as $\frac{TP}{TP+FN}$. Average precision (AP) is computed as the Area Under Curve (AUC) of the precision-recall curve:

$$AP = \int_0^1 p(r)dr \tag{5}$$

When the IoU threshold is 0.50, 0.75, $AP^{0.50}$ and $AP^{0.75}$ are obtained, respectively. AP^{small}, AP^{medium} and AP^{large} are defined according to the pixel value of annotation boxes.

3. The mAP is the mean of AP over all classes given as:

$$mAP = \frac{1}{N} \sum_{i=0}^{N} AP_i \tag{6}$$

4. A weighted score of mAP and IoU is calculated.

$$Score = 0.6 \times mAP + 0.4 \times IoU \tag{7}$$

When Score is same,the standard deviation of AP (Std) is used.

Ablation studies

The convolution is replaced layer by layer from the original backbone network to highlight the contribution of deformable convolution. As shown in Table 2, deformable convolution module improves the model's artefact detection ability, and the detection ability presents an upward trend with the increase of deformable convolution modules. Deformable convolution allows the network to automatically learn the offset of each point during convolution calculation, extract features from the most relevant regions, and focus the convolution area on artefacts as much as possible, avoiding missing detection of small artefacts. When the last three layers are modified to deformable convolutions, compared with base ResNet50, the mAP is increased by approximately 2%, the IoU is increased by 1.44%, and the Score is increased by 1.77%, resulting in optimal detection performances.

Comparison studies

This section will compare the methods employed by the other two EAD2020 challenge teams and the Faster R-CNN to verify the performance of the proposed method in detecting eight artefacts on endoscopic images. This paper's data set division method is consistent with the two teams.

Table 2 Statistical comparisons of ablation studies (%)

Backbone	mAP	IoU	Score
Resnet50-F	40.67	30.17	36.47
Resnet50-FD	41.38	31.93	37.60
Resnet50-FD2	41.23	31.41	37.30
Resnet50-FD3	42.66	31.61	38.24

Table 3 Comparisons with EAD2020 participants (%)

Method	mAP	$AP^{0.50}$	$AP^{0.75}$	AP^{small}	AP^{medium}	AP^{large}
Yu [23]	27.70	53.90	25.00	6.80	12.70	32.30
Chen [24]	26.70	50.10	24.60	8.20	16.20	33.70
Faster R-CNN	35.27	36.80	16.10	18.60	38.80	50.00
Faster RFD3-CNN	42.66	49.10	21.00	20.20	41.10	49.90

Table 3 fully proves that the method we proposed has significantly outperformed others. It increases the mAP by 14.96% and 15.96% higher than the method proposed by [23] and [24], respectively. For artefacts of different scales, the detection results are more accurate, and the detection ability of small artefacts is substantially improved. AP^{small} and AP^{medium} have achieved the best, illustrating that our method has progressed in detecting small and medium artefacts. However, $AP^{0.50}$ and $AP^{0.75}$ are slightly lower. [23] and [24] both adopt Cascade architecture to train detectors using increasing IoU thresholds, which has been proven the gains are mild for low IoU thresholds but significant for the higher ones [14].

In order to prove the effectiveness of the post processing method, tests for four different post processing methods are carried out. Class-agnostic is meant that NMS is applied on all outputs of model, ignoring class difference. The results are shown in Table 4. The best mAP is obtained using class-aware NMS without false positive elimination, which is 44.17%. On the other hand, the post processing method with class-agnostic NMS and false positive elimination achieves the best performance in IoU value, which is 33.22%.

Table 4 Comparisons with various post-processing methods (%)

	Class-agnostic NMS			Class-aware NMS		
Method	mAP	IoU	Score	mAP	IoU	Score
-FP Elimination[1]	38.02	23.08	32.04	44.17	20.09	34.54
+FP Elimination[2]	36.69	33.22	35.30	42.66	31.61	38.24

[1]With false positive elimination.

[2]Without false positive elimination.

Utilizing either class-agnostic NMS or false positive elimination will reduce mAP slightly, but IoU is greatly improved after eliminating some false positives and overlapping boxes. The combination of class-agnostic NMS and false positive elimination leads to a trade-off between mAP and IoU, with a Score of 35.30%. From the perspective of detection performance, the Score with class-aware NMS and false positive elimination is 38.24%, which is the best performance. Our post processing method has a specific contribution to improving detection performance.

Generalization ability

The generalization ability of the artefact detection model on the endoscopic image is defined as the ability of the algorithm to achieve similar performance on the training set when applied to a data set containing the same artefact classes but completely different [2] The 99 endoscopic frames for the out-of-sample generalization task provided by the EAD2020 dataset are used. Table 5 lists test results of the three methods on the generalization ability test set. The method in this paper has relatively superior generalization performance.

Table 5 Comparisons with EAD2020 participants on generalized data (%)

Method	mAP	Std
Chen [24]	24.85	5.52
aAnand [25]	20.20	19.20
Faster RFD3-CNN	24.88	10.12

Overall Performance

Table 6 shows the individual scores for different classes.The Faster RFD3-CNN can successfully detect eight types of artefacts , and the mAP and IoU can reach 42.66% and 31.61%, respectively. Accordingly, the Score is 38.24% (shown in Table 2). Faster RFD3-CNN obtains the best performance with a mAP of 88.4% and an IoU of 76.86% when detecting the instrument in endoscopy images, attributing to metal surfaces and slender shapes of medical instrument greatly differing from the tissues, organs and mucous membranes of the digestive tract. Fig. 3 gives representative prediction outputs from the validation data set and their corresponding ground truth.

Discussion

The ablation, comparison and generalization experiments were designed to verify the performance of the Faster RFD3-CNN model, and the post-processing method proposed in our study, which showed good performance.

Instrument and low contrast are the two types of artefacts most easily detected. Surgical forceps, scissors, and other endoscopic instruments are

Table 6 Overall performances of the proposed method (%)

Class	AP	IoU	$AP^{0.25}$	$AP^{0.50}$	$AP^{0.75}$
Blur	28.95	17.18	38.22	28.47	18.04
Bubbles	20.79	18.70	35.89	21.75	2.49
Specularity	28.43	31.62	46.66	30.46	5.13
Saturation	48.57	26.40	73.88	54.62	11.55
Contrast	61.02	58.16	73.64	65.82	33.53
Instrument	88.40	76.86	92.04	91.09	75.91
Blood	28.71	6.19	44.81	29.18	7.34
Artefact	36.37	17.80	57.42	37.93	10.28

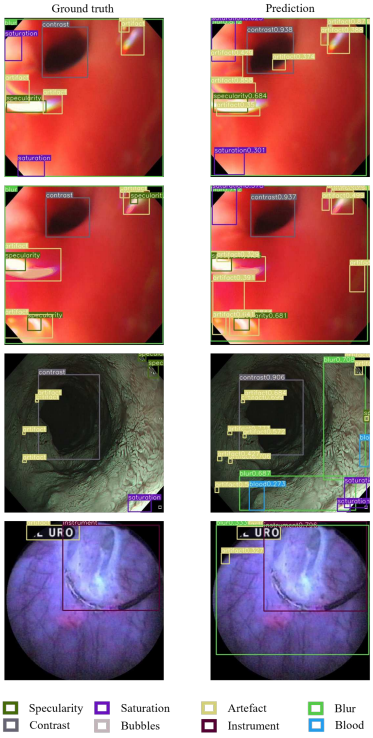

Fig. 3 Representative prediction (right column) and corresponding ground truth (left column)

among the many available; their metal surfaces and slender shapes separate them from the tissues, organs, and mucous membranes found in various digestive tract regions. As a result, instruments are easiest to be distinguished. The low contrast belongs to a common type of artefact with an image proportion of 66.12% (shown in Table 1), mainly because the digestive tract is a hollow pipe. Low contrast features are apparent, often presenting as a circle in the center of the image, with a brightness significantly lower than the surrounding

area, leading to a splendid detecting effect. Saturation and contrast are two forms of artefacts that have opposing qualities. However, because the region of saturation is smaller than that of contrast, and the edge shape is less visible than that of low contrast, both AP and IoU are lower than contrast, with IoU being particularly noticeable. There is a heavy overlap between specular reflections, and each specular reflection region is tiny. Even if there are the most annotations, the main reasons why it is still difficult to improve the detection ability is that specular reflection is easily mistaken for other artefacts. The definition of blur is often subjective to the annotator. The performance of Faster RFD3-CNN in detecting bubbles and blood on endoscopic images needs to be significantly improved. Table 6 presents that our method obtained the worst performance of IoU in detecting blood artefact, which is attributed to the lowest annotation and image proportions and irregular edge shape. The worst AP performance is obtained in detecting bubbles due to their dense distribution and high false detection.

Ablation studies verify the contribution of deformable convolution. The more deformable convolution modules are introduced, the better the detection effect of the method. When comparing the method proposed in this work to others in literature, our method is superior to other methods used in the EAD2020 challenge. At the same time, by comparing the post processing methods, it is found that the significant difference between mAP and IoU is perhaps related to the use of class-aware NMS. If class-agnostic NMS is used, the gap between mAP and IoU will be narrowed at the cost of a drastic reduction in mAP.

The generalization experiment results are still better than the existing methods, but the improvement is slight.

Conclusion

In this study, FPN and deformable convolution are firstly introduced to improve the Faster R-CNN and enhance the detection ability of small object features. Then the Faster RFD3-CNN is proposed to construct the detection model of endoscopy artefacts. Finally, a class-aware NMS and a false positive elimination method based on a confi-dence threshold are used to eliminate false positives as many as possible and get the highest score. Our method can easily predict eight types of artefacts and reach a Score of 38.24% by using lighter networks.

In future work, we hope to solve better the limitation of distribution imbalance to achieve more robust detection. Furthermore, based on this study, we will conduct research on endoscopic image quality assessment and restoration.

Supplementary information.

Declarations

Ethical Approval and Consent to participate. Not applicable.

Human Ethics. Not applicable.

Consent for publication. Not applicable.

Availability of supporting data. Not applicable.

Competing interests. The authors declare no conflict of interest.

Funding. This research was funded by the National Natural Science Foundation (61872405 and 61720106004) and the Medico-Engineering Cooperation Funds from the University of Electronic Science and Technology of China (ZYGX2022YGRH011).

Authors' contributions. Conceptualization, Prince Adjei and Nini Rao; Formal analysis, Tao Gan; Methodology, Han Zhang; Resources, Nini Rao; Software, Han Zhang; Validation, Xiaoming Hu, Xiaotong Wang and Yeting Hu; Writing-original draft, Han Zhang and Nini Rao; Writing-review & editing, Prince Ebenezer Adjei. All authors have read and agreed to the published version of the manuscript.

Acknowledgements. The authors would like to thank the data support from "China Knowledge Centre for Engineering Services and Technology (https://www.ckcest.cn/entry/) and all the reviewers who participated in the review.

Authors' information. The corresponding author is Nini Rao, Ph.D., affiliated to Center for Informational Biology, University of Electronic Science and Technology of China, Chengdu 610054, China and School of Life Science and Technology, University of Electronic Science and Technology of China, Chengdu 610054, China. Tel: +86-028-83208238; Fax: +86-028-83208238; E-mail: raonn@uestc.edu.cn. All information is valid.

References

[1] Ali, S., Zhou, F., Bailey, A., Braden, B., East, J.E., Lu, X., Rittscher, J.: A deep learning framework for quality assessment and restoration in video endoscopy. Medical image analysis **68**, 101900 (2021). https://doi.org/10.1016/j.media.2020.101900

[2] Ali, S., Zhou, F., Braden, B., Bailey, A., Yang, S., Cheng, G., Zhang, P., Li, X., Kayser, M., Soberanis-Mukul, R.D., *et al.*: An objective comparison of detection and segmentation algorithms for artefacts in clinical endoscopy. Scientific reports **10**(1), 1–15 (2020). https://doi.org/10.1038/s41598-020-59413-5

[3] Zhang, R., Zheng, Y., Poon, C.C., Shen, D., Lau, J.Y.: Polyp detection during colonoscopy using a regression-based convolutional neural network with a tracker. Pattern recognition **83**, 209–219 (2018). https://doi.org/10.1016/j.patcog.2018.05.026

[4] Zhang, C., Zhang, N., Wang, D., Cao, Y., Liu, B.: Artifact detection in endoscopic video with deep convolutional neural networks. In: 2020 Second International Conference on Transdisciplinary AI (TransAI), pp. 1–8. IEEE, Irvine (2020). https://doi.org/10.1109/TransAI49837.2020.00007. IEEE Computer Society

[5] Antipov, G., Berrani, S.-A., Ruchaud, N., Dugelay, J.-L.: Learned vs. hand-crafted features for pedestrian gender recognition. In: Proceedings of the 23rd ACM International Conference on Multimedia, pp. 1263–1266. ACM, Brisbane Australia (2015). https://doi.org/10.1145/2733373.2806332

[6] Ojala, T., Pietikainen, M., Harwood, D.: Performance evaluation of texture measures with classification based on kullback discrimination of distributions. In: Proceedings of 12th International Conference on Pattern Recognition, vol. 1, pp. 582–585. IEEE Comput. Soc. Press, Jerusalem, Israel (1994). https://doi.org/10.1109/ICPR.1994.576366

[7] Ojala, T., Pietikäinen, M., Harwood, D.: A comparative study of texture measures with classification based on featured distributions. Pattern recognition **29**(1), 51–59 (1996). https://doi.org/10.1016/0031-3203(95)00067-4

[8] Lowe, D.G.: Object recognition from local scale-invariant features. In: Proceedings of the Seventh IEEE International Conference on Computer Vision, vol. 2, pp. 1150–1157. IEEE, Kerkyra, Greece (1999). https://doi.org/10.1109/ICCV.1999.790410

[9] Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), vol. 1, pp. 886–893. IEEE, San Diego, CA, USA (2005). https://doi.org/10.1109/CVPR.2005.177

[10] Queiroz, F., Ren, T.I.: Endoscopy image restoration: A study of the kernel estimation from specular highlights. Digital Signal Processing **88**, 53–65 (2019). https://doi.org/10.1016/j.dsp.2019.01.012

[11] Segui, S., Drozdzal, M., Vilarino, F., Malagelada, C., Azpiroz, F., Radeva, P., Vitria, J.: Categorization and segmentation of intestinal content frames for wireless capsule endoscopy. IEEE Transactions on Information Technology in Biomedicine **16**(6), 1341–1352 (2012). https://doi.org/10.1109/TITB.2012.2221472

[12] Zhao, Q., Mullin, G.E., Meng, M.Q.-H., Dassopoulos, T., Kumar, R.: A general framework for wireless capsule endoscopy study synopsis. Computerized Medical Imaging and Graphics **41**, 108–116 (2015). https://doi.org/10.1016/j.compmedimag.2014.05.011

[13] Yang, S., Cheng, G.: Endoscopic artefact detection and segmentation with deep convolutional neural network. In: Proceedings of the 1st International Workshop and Challenge on Computer Vision in Endoscopy (2019)

[14] Cai, Z., Vasconcelos, N.: Cascade r-cnn: Delving into high quality object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 6154–6162 (2018). https://doi.org/10.1109/CVPR.2018.00644

[15] Polat, G., Sen, D., Inci, A., Temizel, A.: Endoscopic artefact detection with ensemble of deep neural networks and false positive elimination. In: EndoCV@ ISBI, pp. 8–12 (2020)

[16] Ren, S., He, K., Girshick, R., Sun, J.: Faster r-cnn: Towards real-time object detection with region proposal networks. Advances in neural information processing systems **28** (2015). https://doi.org/10.1109/TPAMI.2016.2577031

[17] Lin, T.-Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2980–2988 (2017). https://doi.org/10.1109/TPAMI.2018.2858826

[18] Artunc, F., Oksuz, I.: An ensemble approach for automatic artefact detection on gastroendoscopy images. In: 2021 6th International Conference on Computer Science and Engineering (UBMK), pp. 741–746. IEEE, Ankara (2021). https://doi.org/10.1109/UBMK52708.2021.9558919

[19] Dai, J., Qi, H., Xiong, Y., Li, Y., Zhang, G., Hu, H., Wei, Y.: Deformable convolutional networks. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 764–773 (2017). https://doi.org/10.1109/ICCV.2017.89

[20] Ali, S., Zhou, F., Daul, C., Braden, B., Bailey, A., Realdon, S., East, J., Wagnieres, G., Loschenov, V., Grisan, E., *et al.*: Endoscopy artifact detection (ead 2019) challenge dataset. arXiv preprint arXiv:1905.03209 (2019). https://doi.org/10.17632/C7FJBXCGJ9.1

[21] Ali, S., Dmitrieva, M., Ghatwary, N., Bano, S., Polat, G., Temizel, A., Krenzer, A., Hekalo, A., Guo, Y.B., Matuszewski, B., *et al.*: Deep learning for detection and segmentation of artefact and disease instances in gastrointestinal endoscopy. Medical image analysis **70**, 102002 (2021). https://doi.org/10.1016/j.media.2021.102002

[22] Huang, Y., Qi, P., Sun, Y.: A novel fault detection algorithm based on the single indicator data. In: 2022 IEEE 25th International Conference

on Computer Supported Cooperative Work in Design (CSCWD), pp. 752–757. IEEE, ??? (2022). https://doi.org/10.1109/CSCWD54268.2022.9776156

[23] Yu, Z., Guo, Y.: Endoscopic artefact detection using cascade r-cnn based model. In: EndoCV@ ISBI, pp. 42–46 (2020)

[24] Chen, H., Lian, C., Wang, L.: Endoscopy artefact detection and segmentation using deep convolutional neural network. In: EndoCV@ ISBI, pp. 37–41 (2020)

[25] Subramanian, A., Srivatsan, K.: Exploring deep learning based approaches for endoscopic artefact detection and segmentation. In: EndoCV@ ISBI, pp. 51–56 (2020)