Improving Loss Function for Polyp Detection Problem

Anh Tuan Tran(✉), Doan Sang Thai, Bao Anh Trinh, Bao Ngoc Vi, and Ly Vu

Le Quy Don Technical University, Hanoi, Vietnam
tuantva86@gmail.com

Abstract. The utilization of automatic polyp detection during endoscopy procedures has been shown to be highly advantageous by decreasing the rate of missed detection by endoscopists. In this paper, we propose a new loss function for training an object detector based on the EfficientDet architecture to detect polyp areas in endoscopic images. The proposed loss combines the features of the Focal loss and DIoU (Distance Intersection over Union) loss named as Focal-DIoU. In addition, we have also carried out some experiments to evaluate the proposed loss function. The experimental results show that our proposed model achieves higher accuracy than previous works on two public datasets.

Keywords: deep learning · EfficientDet · Focal · DIoU · polyp detection

1 Introduction

Polyps are abnormal growths that can develop in various parts of the body, including the colon, stomach, and uterus. In the context of colon health, polyps can potentially develop into cancerous tumors over time [1]. The detection and removal of colorectal polyps are widely regarded as being most effectively accomplished through an endoscopy procedure. However, polyps can be missed in an endoscopy procedure for several reasons, such as, small size of polyps, low experience of doctors and procedure speed [2]. Therefore, computer-aided systems possess significant potential by reducing missed detection rate of polyps to prevent the development of colon cancer [1,2]. Particularly deep learning approaches have shown promising results in improving the accuracy and efficiency of detecting polyps in medical images.

Deep learning models, such as convolutional neural networks (CNNs), have been widely used in medical image analysis due to their ability to automatically learn meaningful features from raw data [12]. By training a CNN-based object detection model on a large dataset of medical images, the model can learn to detect and localize polyps accurately and efficiently. Object detection problems using deep learning approaches such as Regions with CNN features (R-CNN) [15], Fast R-CNN [16], Feature Pyramid Networks (FPN) [20], Single Shot Detector (SSD) [17], You Only Look Once (YOLO) [19], and EfficientDet [27] have gained popularity in recent years. Building upon this achievement,

N. T. Nguyen et al. (Eds.): ACIIDS 2023, LNAI 13996, pp. 209–220, 2023.
https://doi.org/10.1007/978-981-99-5837-5_18

deep learning models have found extensive application in the domain of medical image analysis [12]. Between them, EfficientDet is one of the most effective models for object detection [27].

The loss function for bounding box regression (BBR) is crucial for object detection, with the ln-norm loss being the most commonly used [29]. In BBR, there exists the imbalance problem in training samples. Specifically, the number of high-quality samples (anchor boxes) with small regression errors is much fewer than low-quality samples (outliers) due to the sparsity of target objects in images, e.g., polyps. This paper introduces a novel loss function, named as Focal-DIoU, which integrates the Focal loss and the DIoU loss. The proposed loss function aims to enhance the accuracy of small objects detection while concurrently addressing imbalances in class distribution. Therefore, the detector with the proposed loss function can work well with polyp detection. Moreover, the proposed loss can be easily incorporated into multiple detection models. The experimental results show that the Focal-DIoU loss can enhance the accuracy of the EfficientDet model for the polyp detection problem.

The contribution of our work is summarized as follows:

1. Introduce the Focal-DIoU loss function that handles the imbalance between foreground and background classes together with enhancing the localization of small objects.
2. Integrate the Focal-DIoU loss to train the EfficientDet model for polyp detection.
3. Conduct the various experiments on two well-known polyp datasets. The results show that our proposed loss can enhance the accuracy for the polyp detection problem.

The rest of this paper is organized as follows. Section 2 we first gather essential reviews about our topic. Section 3 presents the backbone EfficientDet and common BBR loss functions. In Sect. 4, we present the proposed method with the new loss function called Focal-DIoU and the object detection model for polyp detection. The experimental settings are provided in Sect. 5. After that, Sect. 6 presents the experimental results and discussions. Conclusions are discussed in Sect. 7.

2 Related Work

In recent years, there have been a number of studies proposed about deep learning-based approaches to object detection such as R-CNN [3], Fast R-CNN [16], FPN [20], SSD [17], YOLO [19], and EfficientDet [27]. There are two types of object detection models, i.e., a two-stage detector and a one-stage detector [10]. The two-stage detector includes a preprocessing step for generating object detection proposals and a detection step for identifying objects. The one-stage detector has an integrated process containing both above two steps.

Two-stage detectors, e.g., Mask R-CNN [22], consist of two separated modules, i.e., a region proposal network (RPN) [3] and a detection module. The RPN generates object proposals, which are then refined by the detection module. This two-stage approach has shown to achieve better accuracy than one-stage detectors, especially in detecting small objects and handling occlusion. However, the disadvantages of the two-stage framework are the requirement of large resources for computation.

To overcome the above shortcomings, one-stage detectors have been developed recently, e.g., YOLO [19], SSD [17], CenterNet [23] and EfficientDet [27] have a simple and efficient architecture that can detect objects in a single phase. These detectors use a feature pyramid to detect objects at different scales and employ anchor boxes to handle object variability. However, they often suffer from lower accuracy compared to two-stage networks, especially in detecting small objects and handling class imbalance. Recently, EfficientDet is one of the most effective object detection model due to using a compound scaling method to balance the model's depth, width, and resolution [24].

Different types of networks can be applied in medical object detection. An object detection algorithm could detect lesions automatically and assist diagnosis during the process of endoscopic examination. Hirasawa et al. [5] used SSD to diagnose the gastric cancer in chromoendoscopic images. The training dataset consisted of 13,584 images and the test dataset included 2,296 images from 77 gastric lesions in 69 patients. The SSD performed well to extract suspicious lesions and evaluate early gastric cancer. Wu et al. [7] proposed an object detection model named ENDOANGEL for real-time gastrointestinal endoscopic examination. ENDOANGEL has been utilized in many hospitals in China for assisting clinical diagnosis. Gao et al. [6] analyzed perigastric metastatic lymph nodes of computerized tomography (CT) images using Faster R-CNN. The results showed that the Faster R-CNN model has high judgment effectiveness and recognition accuracy for CT diagnosis of perigastric metastatic lymph nodes.

The loss function for bounding box regression (BBR) is crucial for object detection, with the ln-norm loss being the most commonly used [29]. However, it is not customized to adapt to the intersection over union (IoU) evaluation metric. The IoU loss is also used in the object detection models. However, the IoU loss will always be zero when two bounding boxes have no intersection [13]. Thus, the generalized IoU (GIoU) loss [25] was proposed to address the weaknesses of the IoU loss, i.e., Recently, the Complete IoU Los (CIoU) and the distance IoU (DIoU) [26] were proposed with faster convergence speed and better performance. However, above losses seem less effective with the imbalance training data. To handle this, we propose the new loss function that combines the IoU based loss, i.e., DIoU and the Focal loss [21] for the polyp detection problem.

3 Background

In this section, we first review the detector used in this paper, i.e., EfficientDet backbone. Second, we describe the mathematical computation of some related loss functions.

3.1 EfficientDet Architecture

EfficientDet [27] is a recent object detection architecture proposed by Tan et al. in 2020. It is based on the EfficientNet backbone [24], which is a family of efficient convolutional neural networks that achieves a state-of-the-art performance on image classification tasks.

One of the key features of EfficientNet is its use of a compound scaling method to balance the model's depth, width, and resolution. This method allows the model to achieve high accuracy with fewer parameters compared to other object detection architectures [24]. Based on EffiecientNet, EfficientDet can extract features from input images effectively. Moreover, EfficientDet employs the Bidirectional Feature Pyramid Network (BiFPN) module [27] to integrate features from multiple scales to improve the accuracy of the detection results. EfficientDet has different versions labeled from D0 to D7 with increasing depth, width, and resolution. In this paper, we use the EfficientDet-D0 backbone which is the smallest and fastest version.

3.2 Loss Function

The loss function is an important component of an object detection model. It helps to guide the training process to enhance the accuracy for the object detection problem. Here, we introduce some common loss functions for the object detection problem.

IoU Loss: The Intersection over Union (IoU) is commonly used as a metric for evaluating the performance of object detection models. It can also be used as a loss function to optimize the model during training [18]. The IoU loss measures the similarity between the predicted bounding box and the ground truth bounding box. It is defined as follows:

$$L_{\text{IoU}} = 1 - IoU, \tag{1}$$

where IoU is the intersection over union between the predicted bounding box and the ground truth bounding box. However, the IoU loss has some weakness when measuring the similarity between two bounding boxes. It does not reflect the closeness between the bounding boxes correctly [13].

Smooth L1: The Smooth L1 loss was first proposed for training Fast R-CNN [16]. The Smooth L1 loss function is defined as follows:

$$L_{\text{Smooth L1}}(x) = \begin{cases} 0.5x^2, & \text{if } |x| < 1 \\ |x| - 0.5, & \text{otherwise} \end{cases} \tag{2}$$

where x is the difference between the predicted and the ground truth bounding boxes.

This loss function is widely used in popular object detection frameworks such as Faster R-CNN and Mask R-CNN because of the smoothness and robustness to outliers. However, similar to the IoU loss, it also do not consider the distance between the bounding boxes. Moreover, the Smooth L1 loss bias to larger bounding boxes.

DIoU Loss: The Distance-IoU (DIoU) loss [26] is proposed to directly minimize the normalized distance between predicted and ground truth bounding for achieving faster convergence. The DIoU loss takes into account the aspect ratio and diagonal distance of the predicted and ground truth bounding boxes. The loss penalizes the distance between the center points of the predicted and ground truth boxes as well as the difference between their diagonal lengths. The formula for DIoU loss can be represented as follows:

$$L_{\text{DIoU}}(b, b^{gt}) = 1 - \text{IoU} + \frac{d(b, b^{gt})^2}{c^2}, \tag{3}$$

where b and b^{gt} denote the central points the predicted and ground truth bounding boxes, respectively; $d(.)$ is the Euclidean distance and c is the diagonal length of the smallest enclosing box covering the bounding boxes. The DIoU loss has shown to be effective to improve the accuracy of object detection models, especially with small objects or many objects in one image [26].

Focal Loss: The Focal Loss is designed to address the imbalance between foreground and background classes during training of object detection models [21]. This loss tries to down-weight easy samples and thus focus training on negative samples. Let's define p is the probability estimated by the model for positive class. Then, we define $p_t = p$ for the positive class and $p_t = 1 - p$ for the negative class. The computation of this loss is as follows:

$$L_{\text{Focal}}(p_t) = -(1 - p_t)^{\gamma} log(p_t), \tag{4}$$

where γ is the focusing parameter that smoothly adjusts the rate for down-weighting easy samples.

4 Methods

In this section, we present the new loss function named as Focal-DIoU to improve the performance of the DIoU loss. After that, we present the polyp detection model with the EfficientNet-B0 backbone trained by the proposed loss.

4.1 The Proposed Loss

The existence of BBR losses has some drawbacks when applying to the polyp detection problem. Firstly, inspired by the improvement of convergence speed as

well as the performance, the IoU-based losses, such as IoU and DIoU, still do not solve the imbalance between high-quality and low-quality anchor boxes. In other hand, the other losses based on the Focal loss are successful in tackling the imbalance problem by increasing the contribution of high-quality boxes [13]. However, these losses only work well with medium or large objects which are not suitable for polyp detection. Therefore, to tackle these above problems, we propose the Focal-DIoU loss which combines the advantages of the Focal and DIoU loss to provide a more effective and robust loss function for training polyp detection models.

Firstly, the Focal-DIoU loss integrates the Focal loss, which is originally designed for addressing the issue of class imbalance in object detection. The idea behind the Focal loss is that it assigns different weights to different samples. Specifically, higher weights are assigned to samples that are miss-classified or hard to classify. This allows the Focal-DIoU to focus more on challenging samples, such as rare objects with larger errors, leading to better optimization and improved performance on imbalanced data.

Secondly, the Focal-DIoU loss also retains the advantages of DIoU, which considers both the aspect ratio and the distance between bounding boxes. By incorporating the distance penalty term, Focal-DIoU can effectively handle variations in object size and position, making it more robust to difference of object scales and object misalignment.

We integrate the DIoU and Focal loss by re-weighting DIoU by the value of IoU, then the Focal-DIoU is computed as below:

$$L_{\text{Focal-DIoU}} = -(1 - \text{IoU})^{\gamma} log(\text{IoU}) L_{\text{DIoU}}. \tag{5}$$

In this paper, we use the Focal-DIoU loss in Eq. 5 to train the detector for the polyp detection problem. The modulating factor $(1 - \text{IoU})^{\gamma}$ intuitively decreases the loss contribution from easy samples and expands the range in which a sample obtains a low loss.

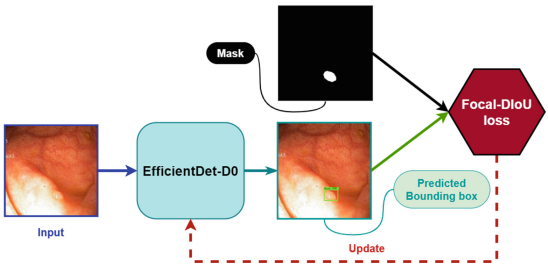


Fig. 1. Architecture of Proposed Model for Polyp Detection.

4.2 Proposed Model

As mentioned in Sect. 3.1, our proposed model is based on the EfficientDet-D0 model as introduced in Sect. 3. The input image is first resized to a fixed size (e.g., $512 \times 512 \times 3$). Then, it is passed through a backbone network EfficientNet-D0, which consists of multiple stages with different spatial resolutions. The backbone network is responsible for extracting multi-scale features from the input image.

As shown in Fig. 1, after passing through the backbone network, the output of the network is a prediction mask. The Focal-DIoU loss function is calculated based on the prediction mask and the ground truth mask with all input samples of a training batch size of the dataset. The value of loss function is used to optimize the weights of the EfficientDet-D0 network.

5 Experimental Settings

This section presents the datasets and the experimental settings used in this paper.

5.1 Datasets

This section presents two polyp datasets used in our experiments, i.e., Kvasir-SEG and CVC-ClinicDB dataset [30].

The Kvasir-SEG dataset [28] is the collection of 1,000 endoscopic images of the gastrointestinal tract, including the esophagus, stomach, duodenum, and colon, obtained from two Norwegian medical centers. The dataset contains images of different types of abnormalities, such as polyps, compression, bleeding, swelling, inflammation, white stool, tumors, and ulcers, with the resolution of $512 \times 512 \times 3$. The image samples are annotated by experts to indicate the location and type of abnormality present in each image.

The CVC-ClinicDB dataset [4] comprises 612 endoscopic images of the colon obtained from the Clinic Hospital of Barcelona in Spain. The dataset includes images of polyps, respectively, acquired using a linear endoscope and a convex endoscope. The images are annotated by experts to indicate the presence or absence of polyps.

In order to prepare the datasets for evaluating the proposed object detection model, we split the datasets into three subsets, i.e., training, validating, and testing, by the ratio as 8:1:1, respectively. The numbers of samples for training, testing, and validating are shown in Table 1.

We transform the mask images of the datasets into bounding boxes to fit with an object detection problem. Here, we apply a contour detection algorithm to the binary mask images to identify the boundaries of the objects in the images. The contours are then converted into rectangular bounding boxes that enclose the objects. This process is executed for each mask image in the datasets. The resulting bounding box annotations are used to train and evaluate our proposed object detection model.

Table 1. Dataset splitting.

Dataset	Total	Train	Test	Val
Kvasir-SEG	1000	800	100	100
CVC-ClinicDB	612	489	61	62

5.2 Parameter Settings

For each dataset, we employ two steps in the training phase. In the first step, the pre-trained EfficientDet-D0 model on the COCO dataset [14] is trained only on the last layer of the EfficientDet-D0 while freezing the rest layers. In the second step, we train all layers of the EfficientDet-D0 model and the early stopping is used to terminate the training process. In the first step, the learning rate, batch size, and the number of epochs are 0.005, 32, and 10, respectively. These values for the second step are 0.001, 16, and 200, respectively.

5.3 Experiment Setup

We conduct the experiments on a computing system with the following specifications: CPU Intel(R) Xeon(R) CPU@ 2.00 GHz, 16 GB of RAM, and a Tesla T4 GPU with 16 GB of VRAM. We use the Python programming language with the PyTorch library [8] to implement our proposed polyp detection model.

For comparison, we train the detectors with the same backbone, i.e., EfficientDet-D0 but using five different loss functions, i.e., Smooth L1 [9], IoU [18], CIoU [11], DIoU [26], and the proposed loss Focal-DIoU. These experiments are conducted on two different datasets, as mentioned in Sect. 5.1. We use COCO metrics [14] for evaluation in our experiments that are based on Average Precision (AP). Notice that, AP is averaged over all classes and we make no distinction between AP and mAP. AP is a performance evaluation metric used in object detection and recognition tasks. It calculates the model's accuracy in determining the location and classifying objects in an image. AR (Average Recall) is a similar performance evaluation metric that focuses on the model's coverage, i.e., its ability to detect all objects present in an image.

6 Results

6.1 Accuracy Comparison

As can be seen from Table 2, the detector with the Focal-DIoU loss achieves the highest AP as 0.637, with competitive performance in other metrics as well. The detector with Focal-DIoU generally outperforms those with the Smooth L1, IoU, CIoU, and DIoU loss in most of the evaluated metrics. On this dataset, the detector with IoU shows the worst performance. Notably, the detector with DIoU achieves the highest values in some specific metrics such as AP_{75} and

AP_L. However, the detector with Focal-DIoU remains the top-performing loss function in terms of overall performance with the highest AP score. The reason is that the proposed loss function helps the training detector by considering the small, medium, and the large size polyps. Thus, using our proposed loss function enhances the overall accuracy compared with the previous loss function, such as Smooth L1, IoU, CIoU, and DIoU on the Kvasir-SEG dataset for the polyp detection problem.

Table 2. The results of EfficientDet-D0 with different loss functions on Kvasir-SEG dataset.

LossFunction	*AP*	AP_{50}	AP_{75}	AP_L	*AR*	AR_{50}	AR_{75}	AR_L
Smooth L1	0.612	0.832	0.715	0.695	0.638	0.657	0.657	0.743
IoU	0.542	0.769	0.583	0.617	0.557	0.629	0.633	0.709
CIoU	0.619	**0.853**	0.709	0.702	0.637	0.672	0.672	0.755
DIoU	0.623	0.848	**0.727**	0.704	0.645	**0.691**	**0.691**	**0.772**
Focal-DIoU	**0.637**	0.850	0.680	**0.721**	**0.663**	0.679	0.679	0.764

Similarly, as is presented in Table 3, the detector with the Focal-DIoU loss achieves the highest *AP* as 0.790. The detector with the CIoU loss also delivers competitive results with the highest Recall as 50% IoU threshold (i.e., AP_{50}). Generally, the detectors with the Focal-DIoU and CIoU loss outperform others in almost all of the evaluated metrics. We can observe that Focal-DIoU achieves the highest values in most metrics except AR_{50} and AR_{50} where CIoU gets the highest score. Overall, the proposed loss function helps the detector for improving the accuracy for the polyp detection problem.

Table 3. The results of EfficientDet-D0 with different loss functions on CVC-ClinicDB dataset.

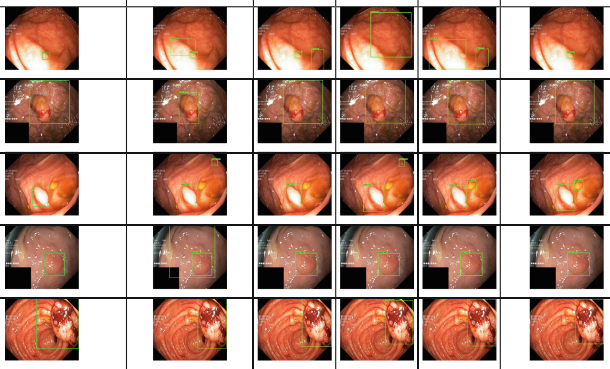
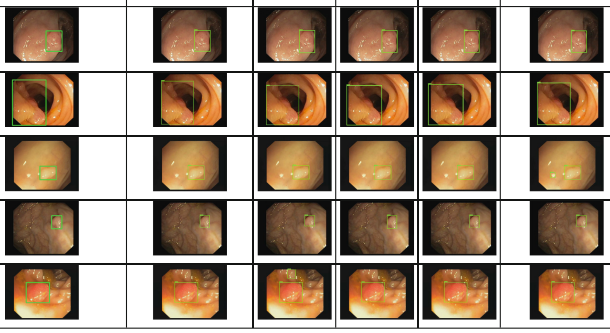
LossFunction	*AP*	AP_{50}	AP_{75}	AP_L	*AR*	AR_{50}	AR_{75}	AR_L
Smooth L1	0.773	0.993	0.879	0.753	0.787	0.829	0.829	0.817
IoU	0.773	0.957	0.909	0.773	0.783	0.808	0.808	0.813
CIoU	0.789	0.986	0.893	0.778	0.806	**0.835**	**0.835**	0.830
DIoU	0.776	0.949	0.886	0.767	0.790	0.808	0.808	0.803
Focal-DIoU	**0.790**	**0.996**	**0.926**	**0.805**	**0.814**	0.827	0.827	**0.840**

6.2 Visualization

To observe the results of the polyp detection visually, we show the detection results for several images of the experimental datasets in Table 4. This table

shows that the detector with the Focal-DIoU loss helps to detect the polyp area more correctly in both experimental datasets. Especially, on the Kvasir-SEG dataset, the detector with the Focal-DIoU loss achieves more correctly polyp detection results compared with the detectors with other loss function. For the CVC-ClinicDB dataset, the detectors with loss functions achieves similar accuracy. Overall, the detector with the Focal-DIoU loss presents the best detection result even with very small and unevenly distributed polyps in the image.

Table 4. Visualization of polyp detection resulting from EfficientDet-D0 with different loss functions.

Dataset \ Loss func.	Original Image	Smooth L1	IoU	CIoU	DIoU	Focal-DIoU
Kvasir-SEG						
CVC-ClinicDB						

7 Summary

In this paper, we focus on studying loss functions to improve the accuracy and efficiency of the polyp detection model. We propose Focal-DIoU loss to train the effective detector, i.e., EfficientDet-D0 backbone for the polyp detection problem. The proposed loss function can help the detector consider small, medium, and large size of polyps in the training process. Thus, the detector enhances the

accuracy to detect various sizes of polyps. The experimental results show that the detector with the proposed loss function achieves higher accuracy than the detectors with other loss functions for the polyp detection problem. This proves that the proposed loss function can help to improve the polyp detection problem.

References

1. Lee, S.H., et al.: An adequate level of training for technical competence in screening and diagnostic colonoscopy: a prospective multicenter evaluation of the learning curve. Gastrointest. Endosc. **67**(4), 683–689 (2008)
2. Leufkens, A., Van Oijen, M., Vleggaar, F., Siersema, P.: Factors influencing the miss rate of polyps in a back-to-back colonoscopy study. Endoscopy **44**(05), 470–475 (2012)
3. Girshick, R., Donahue, J., Darrell, T., Malik, J.: Region-based convolutional networks for accurate object detection and segmentation. IEEE Trans. Pattern Anal. Mach. Intell. **38**(1), 142–158 (2015)
4. Bernal, J., Sánchez, F.J., Fernández-Esparrach, G., Gil, D., Rodríguez, C., Vilariño, F.: WM-DOVA maps for accurate polyp highlighting in colonoscopy: validation vs. saliency maps from physicians. Comput. Med. Imaging Graph. **43**, 99–111 (2015)
5. Hirasawa, T., et al.: Application of artificial intelligence using a convolutional neural network for detecting gastric cancer in endoscopic images. Gastric Cancer **21**(4), 653–660 (2018). https://doi.org/10.1007/s10120-018-0793-2
6. Gao, Y., et al.: Deep neural network-assisted computed tomography diagnosis of metastatic lymph nodes from gastric cancer. Chin. Med. J. **132**(23), 2804–2811 (2019)
7. Wu, L., et al.: A deep neural network improves endoscopic detection of early gastric cancer without blind spots. Endoscopy **51**(06), 522–531 (2019)
8. Paszke, A., et al.: Pytorch: an imperative style, high-performance deep learning library. In: Advances in Neural Information Processing Systems, vol. 32 (2019)
9. Fu, C.Y., Shvets, M., Berg, A.C.: Retinamask: learning to predict masks improves state-of-the-art single-shot detection for free. arXiv preprint arXiv:1901.03353 (2019)
10. Du, L., Zhang, R., Wang, X.: Overview of two-stage object detection algorithms. J. Phys. Conf. Ser. **1544**(1), 012033 (2020)
11. Wang, X., Song, J.: ICIoU: improved loss based on complete intersection over union for bounding box regression. IEEE Access **9**, 105686–105695 (2021)
12. Puttagunta, M., Ravi, S.: Medical image analysis based on deep learning approach. Multimedia Tools Appl. **80**(16), 24365–24398 (2021). https://doi.org/10.1007/s11042-021-10707-4
13. Zhang, Y.F., Ren, W., Zhang, Z., Jia, Z., Wang, L., Tan, T.: Focal and efficient IOU loss for accurate bounding box regression. Neurocomputing **506**, 146–157 (2022)
14. Lin, T.-Y., et al.: Microsoft COCO: common objects in context. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014. LNCS, vol. 8693, pp. 740–755. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10602-1_48
15. Girshick, R., Donahue, J., Darrell, T., Malik, J.: Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 580–587 (2014)

16. Girshick, R.: Fast R-CNN. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1440–1448 (2015)
17. Liu, W., et al.: SSD: single shot MultiBox detector. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9905, pp. 21–37. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46448-0_2
18. Yu, J., Jiang, Y., Wang, Z., Cao, Z., Huang, T.: Unitbox: an advanced object detection network. In: Proceedings of the 24th ACM International Conference on Multimedia, pp. 516–520 (2016)
19. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: unified, real-time object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 779–788 (2016)
20. Lin, T.Y., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S.: Feature pyramid networks for object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2117–2125 (2017)
21. 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)
22. He, K., Gkioxari, G., Dollár, P., Girshick, R.: Mask R-CNN. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2961–2969 (2017)
23. Duan, K., Bai, S., Xie, L., Qi, H., Huang, Q., Tian, Q.: Centernet: keypoint triplets for object detection. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 6569–6578 (2019)
24. Tan, M., Le, Q.: Efficientnet: rethinking model scaling for convolutional neural networks. In: International Conference on Machine Learning, pp. 6105–6114. PMLR (2019)
25. Rezatofighi, H., Tsoi, N., Gwak, J., Sadeghian, A., Reid, I., Savarese, S.: Generalized intersection over union: a metric and a loss for bounding box regression. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 658–666 (2019)
26. Zheng, Z., Wang, P., Liu, W., Li, J., Ye, R., Ren, D.: Distance-IoU loss: faster and better learning for bounding box regression. In: Proceedings of the AAAI Conference on Artificial Intelligence, pp. 12993–13000 (2020)
27. Tan, M., Pang, R., Le, Q.V.: Efficientdet: scalable and efficient object detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 10781–10790 (2020)
28. Jha, D., et al.: Kvasir-SEG: a segmented polyp dataset. In: Ro, Y.M., et al. (eds.) MMM 2020. LNCS, vol. 11962, pp. 451–462. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-37734-2_37
29. Wang, Q., Cheng, J.: LCornerIoU: an improved IoU-based loss function for accurate bounding box regression. In: 2021 International Conference on Intelligent Computing, Automation and Systems (ICICAS), pp. 377–383. IEEE (2021)
30. Wang, F., Hong, W.: Polyp dataset (2022). https://doi.org/10.6084/m9.figshare.21221579.v2