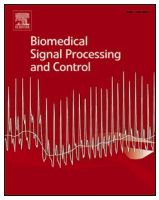
Biomedical Signal Processing and Control 80 (2023) 104283



Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

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





Colon cancer stage detection in colonoscopy images using YOLOv3 MSF deep learning architecture

Malathi Murugesan [a,*], R. Madonna Arieth [b], Shankarlal Balraj [c], R. Nirmala [d]

[a] Department of Bio-Medical Engineering, E.G.S. Pillay Engineering College, Nagapattinam, Tamil Nadu 611002, India
[b] Department of CSE (AI & ML), Sreenivasa Institute of Technology And Management Studies (SITAMS), Chittoor, Andra Pradesh, India
[c] Department of ECE, Perunthalaivar Kamarajar Institute of Engineering and Technology, Karaikal, Puducherry 609603, India
[d] Vivekanandha College of Engineering for Women (Autonomous), Elayampalayam, Thiruchengode, Namakkal 637205, Tamilnadu, India

A B S T R A C T

Colo-Rectal Cancer (CRC) stood witnessed as the major cause of deaths worldwide, especially in men. But the early detection and removal of polyps can reduce the death rate to a considerable extent. The various stages of colon cancer using the YOLOv3 Multi-Scale Framework (YOLOv3-MSF) have been studied here in this context. Colonoscopy is an examination method to detect changes in the rectum by inserting a long flexible tube into the rectum. For screening colon cancer, gastroenterologist often refers to the pathological images of colonoscopy for diagnosis. In recent decades, various machine learning algorithms are used for diagnosis to get practical and reliable results. More often, the stages of colon cancer can be inspected by the Tumor Length (TL). In this study, the polyps are segmented and allowed to measure the TL value by assessing polyp length (P_{len}) and polyp width (P_{wid}). The various stages of colon cancer have been ascertained based on the maximum length between the P_{len} and P_{wid}. Our research uses the single-stage YOLOv3 object detection model to locate and classify the various stages based on the TL value. It outputs the class probability and position coordinates of each image by enhancing accuracy and robustness. At the outset, during preprocessing to address the specific problems like reduced contrast, blurred images and noisy data, the proposed model defines a data augmentation with a jitter value of 0.3 to avoid overfitting. Next, for detecting and annotating various stages of colon cancer, YOLOv3-MSF deep learning architecture with multiscale detection layers used here. By adopting transfer learning, we used YOLOv3 in our work. It chooses the anchor boxes using the K-Medoids algorithm during object detection in the training stage based on the ResNet network structure of YOLOv3. Finally, additional layers such as fully connected layer are embedded into the existing architecture. we used the CVC colonDB database to train and evaluate the fully connected layer during experiments. To analyze the performance of the proposed system, we compare our model with the state of art object detection model. Experimental results show that the proposed model outperforms existing systems on peak to accuracy, precision rate, recall rate, and f-measures.

1. Introduction:

Colorectal cancer is the leading cause of death in the world [1]. Early detection and treatment of colorectal cancer will result from effective colorectal cancer screening, lowering medical expenses. Adenoma of the large intestine is the source of colorectal cancer tumors [2]. Its is considerably very difficult to recognize and recover adenoma occur in the large intestine or cancer at the early stage without putting people through unnecessary tests or stress [3]. If adenoma occur in the large intestine, otherwise colorectal cancer is anticipated, colon endoscopy will be conducted qualitatively analyzed. Recently, various hospitals ready doing colonoscopy magnification of colon endoscopies in addition with enhancement of image methodologies includes narrow band imaging (NBI), furthermore examination test became unavailable [4–6]. According to Japanese Society of 2016, treatment which exist in Colon and Rectal cancer treatment guidelines of cancer which present in polyps, adenoma, CT is cancer and invasive cTI cancer (hereinafter, cT1a), single resection through colon endoscopy is possible. However, there is a 10 %–15 % or more chance of lymph vascular invasion in profoundly aggressive cT1 (SM) (hence, cT1b) or throughout deepened colorectal cancer, necessitating surgical resection with lymph node dissection [7–9]. Avoiding needless endoscopic removal during endoscopic diagnosis will result in performing direct surgical resection, which provides costs and time advantages [10].

In most endoscopic diagnoses, in order to adjust luminal expansion from the endoscope, in the presence of both aeration and de-aeration is present which is done with help of non-magnified observation [1]. Also, certain things like form changes, distension and tumor elasticity, exist and in-exist of recession and two leveled protrusion and folds of concentration which is observed on the surrounding surface. If the NBI magnification findings are JNET 2B and JNET3, a pit pattern diagnostic utilizing crystal violet staining and white light magnification and

* Corresponding author.
E-mail address: m_malathi123@rediffmail.com (M. Murugesan).

https://doi.org/10.1016/j.bspc.2022.104283
Received 30 September 2021; Received in revised form 14 September 2022; Accepted 26 September 2022
Available online 10 November 2022
1746-8094/© 2022 Elsevier Ltd. All rights reserved.

observation is indicated [11–14]. Diagnosis with magnified observations and NBI observations have been reported to improve diagnostic rates even for tiny and small colorectal polyps that may be difficult to diagnose. Colon endoscopies, on the other hand, have grown in popularity and improving endoscopic diagnosis rates in routine colon endoscopy examinations is crucial. Furthermore, cT1b has determined that under conventional white light observation with 0.1 percent indigo carmine application, the specificity, sensitivity and diagnosis at the proper rate which requires surgical resection are 66.0, 95.8, and 86.3 percent respectively [15]. Medical expertise involves through the functionality performing the endoscopy activity and on the other hand, are under pressure to perform correct depth of invasion diagnosis which is considered the phases of colon endoscopies at the time of screening in clinical environments [16].

Target identification is among the most key concepts in medical image diagnosis, also it has long been a wider research area. Several detectors like deep convolutional neural network (DCNN) [17] have indeed been suggested along with advancement of GPU which is graphical processing unit [GPU] hardware technology. DCNN outperforms classical target detectors in terms of generalization and description. This will be automatically learning the content with respect to semantic of the respective targets. These detectors are still widely employed in our daily lives [18], For instance, recognition of face [20,21], recognition of action [22,23], transportation in intelligence way [24–26], driving uses autonomous track [27,28], surveillance adopted intelligent [29–31], retrieval of clothing purpose for advertising videos [19]. Multi-scale output is employed by detectors that works on the functionality of DCNN, which includes YOLO V3 [32], SSD [33], and DSSD [34], to detect targets of various sizes. With a highest-level pathway including lateral interconnections, by actioned this situation, feature pyramid network (FPN) which is progress by means of combining the feature of two adjacent feature hierarchy scale [35]. Highly fine-grained information that are provided by feature with respect to high resolution of feature map. It is very useful will be able to forecast the position and category of target with highest accuracy. The classification accuracy which is greatly enhanced by implementing the FPN concept in the architecture of YOLO V3 and DSSD.

Despite the fact that detector with respect to multi-scale target detection has been improved the performance of target detection, still there is significant potential which means for development. For starters, feature fusion across adjacent scales isn't enough. It's crucial to combine feature maps with different scales in order to better find and classify the targets. Furthermore, prior to feature fusion, the feature map size must be unified. Also up-sampling and down-sampling which has been utilized to reformat vectors respect to features at different levels to make them the same size. Interpolation is a common method of up-sampling. The ill-posed restoration problem [36] cannot be solved with interpolation because it lacks extra information. Down-sampling approaches have always been max-pooling, average-pooling as well. Pooling, down sampling which leads to some critical information has been losses. It is unable to provide a method for learning the mapping process from high resolution to low resolution. The process of concatenation also utilizing to merge the two features by joining the feature map channels and considerably that is popular feature fusion process. It cannot, however, convey the significance of features which is spatial at different scales also channel features of corresponding correlation.

Attempts to predict outcomes from the clinical results using histopathological evidence which utilizes machine learning have employed either of two strategies. Pre-defined morphologic using proprietary tools like Cell Profiler which is the very first strategy focuses on extraction of features. Then to figure-out pre-defined traits, both machine learning and statistical approaches are linked also correlated to survival. Then the second and very advent technique is to employ insufficient deep learning supervised approaches to predict directly from the survival of WSI. Then removing the need for pre-defined features but adding new hurdles in terms of model explicability. While several researches have attempted to depict the morphological traits acquired mostly by models, repeatable interpretations of these kind of features and determine the level over which they truly explain the prediction of model which constantly maintains the problems. Deep learning is a type of machine learning algorithms that uses increasingly abstract representations of incoming data to execute a job. These methods make use of training data to figure out how to build these representations in a task-appropriate way. Conventional machine learning, on the other hand, creates representations of the input data that are used to execute the task using handcrafted features. Deep learning has already been proved to outperform previous machine learning approaches in a variety of applications, and it is predicted to improve clinical procedure. Convolutional neural networks have performed at a variety of picture interpretation tasks, and it's possible that they may be used to extract additional data from histopathological photos. The goal of this study was to apply deep learning to analyses traditional whole-slide photographs in order to generate an automatic predictive biomarker for patients with primary colorectal cancer who had their disease removed.

Hence, the objective of the current research focuses, to avoid overfitting problem, data augmentation is utilized. Next, for detecting and annotating various stages of colon cancer, YOLOv3-MSF deep learning architecture with multiscale detection layers used here. Once tumor is detected over the image then the length of tumor is determined. We utilized YOLOv3 architecture for our proposed work with help of pretrained model. Further more fully connected layer is developed in the last layer of existing architecture in which we used the CVC colonDB database for additional training. To analyze the performance of the proposed system, we compare our model with state-of-art, Faster-CNN object detection model.

2. Related study and literature survey:

This section describes various related study that are invented by the researchers so far. This survey only focuses through, how deep learning techniques which are widely utilized in order to detect colon cancer.

Traditional method and convolutional neural network are the two group of composed techniques that are widely used for the detection of polyp. The conventional polyp detection technique which is centered on handcrafted features which is systematically works on the basis of separate classifier. The detection of polyp which uses the geometric features which includes shape, size and information of texture [37–39]. For instance, in the literature work for detecting polyp and geometric and texture features are widely used. In order to find out image contour, an edge detection through canny filter is used [37]. Then upon that way, to facilitate automated selection marker for the region of polyps, clustering technique along with gabor texture features [40] is used. For the identification of polyps, a geometric information has been extracted. Detection of automated polyp segment which is detection by means of utilizing support vector machine and cross section profiles.

CNN has rapidly drawn a huge amount of attention to the field of computer vision but have achieved significant progress in a number of areas, including detection [55–57], segmentation [54] and classification [53]. In addition, CNN has been used in analysis of healthcare sector generated image. For example, the demonstration of deep learning approach can outperform pathologists in terms of diagnostic performance [48]. Also, high level of sensitivity and specificity which is produced by various deep learning algorithms in order to detect tumors.

In addition, some algorithms have used features for recognition of polyp, assuming that polyps get a lot of information regarding texture. For illustrate, the author created classification technique which means histogram feature of two dimension by combining the Local Binary Pattern (LBP) with average intensity [46]. The authors [47] used patches of different size to depict the image portions at the same time. They used the size of multiple patches to train an SVM classifier, that can also spot anomalous regions quite accurately than patches of single size alone.

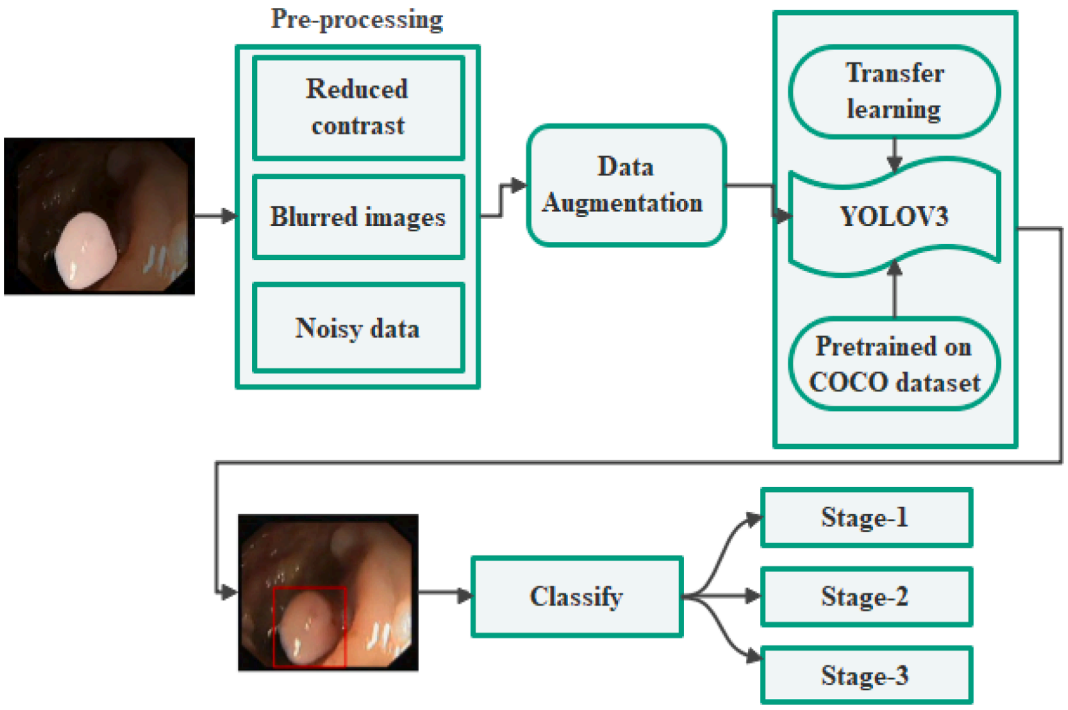


Fig. 1. Proposed Work flow.

For automatic detection of polyp, a two-level technique has been proposed by the author [42]. In the given context, the considered parameters are texture and color of data, features of temporal information, and shape in the given context were used to find probable polyp areas in research literature. Further on, CNN was used to learn as well as classify features. The overall advantage of CNN to make use of entire - information over the given image, which is included by color, shape and texture. For various analysis of CNN, the authors conducted different strategies which resulted on basis of reflecting towards the trained model from the scratch and furthermore, which is fine tuned with help of various types of cancer. They determined that fine-tuned algorithm outperforms well also, when compared with ROC curves, it proves full training by using fine-tuned function.

A three-dimensional deep learning architecture which was developed by Yu et al. [49,50] that may be utilized to minimize the rate of false positive and increase performance of network discrimination. In order to transform low level features from different dataset of non-medical domains which widely utilizes the detection and classification models by Zhang et al. [43,44]. The suggested technique distinguished between from the features of polyp and non-polyp before predicting the morphology of polyps in the latter. This presented scheme shown that in order to utilize transfer learning can be quite effective in medical tasks [45]. A detection approach, which is based on CNN architecture also it includes different down and up sampling techniques [51,52]. This method was unique because of how it investigated different feature of multi-level representations with fully CNNs from beginning to end.

3. Methodologies:

This section elucidates different approaches that are used for our proposed system. Our proposed system utilizes single-stage YOLOv3 object detection model in order to detect and locate the various stage based on the TL value. It outputs the class probability and position coordinates of each image by enhancing accuracy and robustness.

Initially preprocessing is done which includes minimizing contrast over the image, blurred image and noisy data. Once preprocessing is done, then to avoid overfitting problem, data augmentation technique is utilized. Then YOLOv3-MSF model is utilized, which is a pretrained model and trained using COCO dataset. Furthermore, two fully connected layer is applied at the last layer of existing architecture, in order improve the efficiency of the model. We used CVC colonDB database is used in order to train the fully connected layer. Fig. 1, clearly depicts the proposed work flow.

3.1. *Data pre-processing:*

In general, preprocessing is the technique to remove all the noise and unnecessary thing from the given set of datasets. Our collected raw dataset, which is CVC colonDB. This dataset is un preprocessed in nature which means that, it contains the noise among the dataset, also we convert that into desired vector at this stage. The first stage in processing the dataset is preprocessing, which is increasing the quality of an image before it is utilized further by removing undesired image information, often known as image noises. If this issue is not addressed appropriately, there may be several mistakes in the classification. The treatment is also required due to flaws such as low contrast between the lesion and surrounding healthy skin, an uneven border, and skin aberrations such as hairs, skin lines and black frames. Many filters can be used to eliminate gaussian noise, speckle noise, Poisson noise, salt and pepper noise, which includes the filters of means, median, adaptive median, gaussian filter and adaptive wiener filter. For instance, an image consisting of hairs with the lesion, which may lead to misclassification.

Contrast adjustment, vignetting effect removal, color correction, picture smoothing, hair removal, normalizing and localization are all pre-processing techniques that are used to reduce or eliminate image noise. Combining the right pre-processing procedures yields more precision. Preprocessing techniques includes black frame removal, automatic color equalization, hair removal, dull razor, karhunen–loeve transform, gaussian filter, pseudo-random filter, non-skin masking, color space translation and feature extraction.

3.2. *Data Augmentation:*

Simple changes such as horizontal flipping, augmentations of color and space, and random cropping were the first to demonstrate the usefulness of Data Augmentations. Several invariances that pose a difficulty

of recognition of image tasks are encoded via these transformations. Some the widely utilized augmentations technique that people used for their image dataset are, transformations of geometric, transformation of color space, images mixing, augmentation of feature space, training using adversarial network, transfer style of neural and schemes based on *meta*-learning. Here we utilized augmentation technique in order to avoid overfitting problem by using jit value 0.3.

3.3. YOLOv3 Multi-Scale Framework:

Most widely used detection of objects in the recent days is YOLO which was proposed by joseph redmon et al. [41]. But nowadays, it plays a vital role in order to provide promising outcomes over the field of detecting objects. Then YOLOv3 was evolved on the basis of YOLO, which is a well advanced and latest object detecting technique. A single deep convolutional neural network is utilized in YOLOv3, furthermore it divides the given input image into numerous cells and each cell predicts a bounding box directly and classify the objects. The illustration of YOLOv3, which is depicted in figure. The important function of that algorithm is to divide the given image into M × M cell, then prediction is occurred in every cell. Furthermore, middle position of an object is found in every cell, that is used in order to liable each and every identified object. Consider each bounding boxes, a confidence score is determined which is used to predict an all-bounding box. Consider every box, this confidence score is the indicator of an object and it helps to determine the cell. Also, it helps to predict every box accurately. Also, confidence score is calculated by using the given equation,

$$Confidence = Q_r(obj) \times IoU(groundtruth\ value,\ predbox) - - - \quad (1)$$

From equation (1), the range of Q_r from $Q_r(obj) \in \{0,1\}$ which indicates that the intersection over the union in-between ground-truth value (GT) and predicted box (PredBox). Moreover, the object does not exist in the cell which indicates that the range of confidence score should be {1}. In otherwise we can say, both IoU-intersection over union and confidence score is equal which lies between value of ground truth and predicted box. Consider to detecting object, over each predicted class probability which is conditional "C". Thus, for each cell the expected total value is $5 + C$. And the expected value parameters include $\{a,b,w_j, h_j, conf\}$. The coordinates of middle position of the box are indicated by $\{a,b\}$. Then the width and height of the respective boxes is indicated by w_j, h_j. Various issues which are solved by YOLOv3 that includes normalization of batch, problem regarding skip connection and problems regarding localization. Also, YOLOv3 detect object on basis of three scales. This is done with help of input image size up to 32, 16 and 8 respectively, also down sampling technique is utilized to in terms of accurate prediction. The prediction is done with help of encoding process of YOLOv3, which is given in equation (2).

$$N \times N \times (B \times 5 + C) - - - \quad (2)$$

Each input image of cell size is $N \times N$. The total bounding boxes in the cell is 'B', for the attributes of four bounding box is '5′ with respective number of cases is indicated as 'C'. The target object which is consisting of different classes, hence convolutional layer such as 1×1 along with activation function of logistic regression are used instead of using SoftMax layers. By means of using threshold value, a target object is detected also by using logistic regression, a confidence score is projected over each determined class. The error occurred in predicted and real value is computed by using the loss function includes classification, localization and confidence. By combining this three-loss function, a final cross-entropy loss function is determined which is .

Table 1
Process of resizing feature map of every scale.

YOLO V3	Resizing of scale1	Resizing of scale1	Resizing of scale1
Fusion of feature for scale 1	–	Convolution layer as subpixel denotes (r = 2)	Convolution layer as subpixel denotes (r = 4)
Fusion of feature for scale 2	Layer of pathway denotes (r = 2)	–	Convolutional layer as subpixel denotes (r = 2)
Fusion of feature for scale 3	Layer of pathway denotes (r = 4)	Layer of pathway (r = 2)	–

$$\begin{aligned} Lossfunction = {} & \lambda_{cord} \sum_{i=0}^{T^2} \sum_{j=0}^{C} I_{ij}^{obj} \\ & \times \left[\left(\sigma(p_a)_i^j - \sigma(\widehat{p}_a)_i^j\right)^2 + \left(\sigma(p_b)_i^j - \sigma(\widehat{p}_b)_i^j\right)^2 \right] + \lambda_{cord} \sum_{i=0}^{T^2} \\ & \times \sum_{j=0}^{C} I_{ij}^{obj} \times \left[\left(p_{w_i^j} - \widehat{p}_{w_i^j}\right)^2 + \left(p_{h_i^j} - \widehat{p}_{h_i^j}\right)^2 \right] + \sum_{i=0}^{T^2} \\ & \times \sum_{j=0}^{C} G_{ij} \left(B_i^j - \widehat{B}_i^j\right)^2 + \lambda_{nobj} \sum_{i=0}^{T^2} \\ & \times \sum_{j=0}^{C} I_{ij}^{nobj} \left(B_i^j - \widehat{B}_i^j\right)^2 + \lambda_{nobj} \sum_{i=0}^{T^2} I_{ij}^{nobj} \sum_{b \in classes} \left(t_i^j(c) - \widehat{t}_i^j(c)\right)^2 - - - \end{aligned}$$

From equation 3, I_{ij}^{obj} indicates that, in order to predict target object, jth bounding box is responsible for the cell belongs to ith indices. If this gets satisfies the equivalent value ranges as 1, otherwise it is to be 0. The number of cells is indicated as S^2 and number of bounding boxes of each cell is represented as B. The middle position of each cell represented as $\sigma(p_x)$ and $\sigma\left(p_y\right)$ respectively. The target class probability value is represented by 'p' along with respective confidence score is indicated as 'c'.

3.4. Detection of Tumor's region:

Once, bounding box is determined over the image, which represent that all the tumors' patches are identified. With respect to the loss function, a generalized IoU, is performed over the identified boundary box. This can be used find out the predicted bounding box and ground truth object is how match to each other in order to detect the length of tumors using bounding boxes. It i3s computed with help of equation 4,

$$IoU_{T_{GT},T_{GT},} = \frac{|T_{GT} \cap T_p|}{|T_{GT} \cup T_p|} - \frac{\left|T \backslash \left(T_{GT} \cup T_p\right)\right|}{|T|} \quad (4)$$

From equation (4), the ground truth value of bounding box is represented as T_{GT} along with respective predicted bounding box as, T_p. Furthermore, the respective loss is determined with by using the formula,

$$IOU\ loss = 1 - IOU_{Bgt,\ Bp} - - - \quad (5)$$

To predict the confidence of target class is represented with respect to focal loss. It can be reduced with respect to relative along with much attention needed over the detected region. Once this is done then the process resizing feature map is performed which is clearly indicated in Table 1,

3.5. Non-Maximum suppression:

In order to capture each object perfectly, multiple bounding boxes are being created by YOLOv3 while detecting target object over an image which is also be different shapes and different sizes in nature. By

Table 2
Comparison between YOLO-V3(Without transfer learning) with other state of art models.

Model	Accuracy	Precision	Recall	F-Measure
ResNet	81.93 %	79.12 %	78.03 %	78.98 %
AlexNet	83.32 %	80.23 %	79.43 %	80.01 %
Fast-CNN	84.53 %	81.09 %	82.45 %	82.01 %
Faster-CNN	87.87 %	85.23 %	86.39 %	85.98 %
Deep Learning ANN	93.26 %	91.04 %	92 %	86.01 %
YOLO-V3(Without transfer learning)	93.4 %	91.08 %	91.78 %	91.32 %

means of detecting single bounding box from the set of various overlapping bounding boxes, a non-maximum suppression technique is widely used. This technique is the popular technique because its simplest nature. For detection of best match result, this method eliminates detection of redundant objects. The algorithm for non-maximum suppression is given as follow:

Algorithm: Begin.

Input: D $= \{d_1, d_2, \cdots . d_n\}, S = \{s_1, s_2, \cdots . s_n\}, T_h$

D - initial set of detected boxes
S - Respective confidence score
T_h- Threshold for NMS
Output: Set of final detected boxes D
1: B ←{}
2: while D! $= \varnothing$ **do**
3: $m = maxargS$
4: $B \leftarrow B \cup d_m; D \leftarrow D - d_m; S \leftarrow S - m$
5:for $d_i \in Dbo$
6:if $(d_m, d_i) \geq T_h$ then
7: $D \leftarrow D - d_i; S \leftarrow S - i$
8:end **if**
9:end **for**
10: end **while**

4. Experimentation and result discussion:

Our proposed system, which is developed by using YOLOV3 object detection model. We utilized pretrained model of YOLOV3, which is originally trained by using COCO dataset. In order to improve the model accuracy, an additional layer is embedded to existing architecture. We provided additional layer with help of fully connected layer which is trained by using 30 epochs. The dataset that, we are using for additional training is CVC colonDB database. Once detection is done with help of developed model, then on the basis of length of tumor, we classify that belongs to stage 1, stage 2 or stage 3.

4.1. Description of dataset:

Dataset consisting of 1000 HD endoscopic images that are widely collected from 3 clinical procedure rooms. Furthermore, the dataset consists of early cancerous lesions of example as 500 and non-cancerous lesions of example 500. Also, the annotation was independently done with help of 5 top medical experts from Europe, by means of find out appropriate standard. All experts are used a various color for delineation of cancer. The dataset images consist of 598 patients' images, which consisting both male and female. We collected CVC colon dataset from https://endovissub-barrett.grand-challenge.org/Data_description/. This dataset is trained with help of two fully connected layer of 30 epochs.

4.2. Result comparison:

Once, experimentation is done by utilizing with and without transfer learning techniques. Here, we compared results with state of art model such as ResNet, AlexNet, Fast-CNN and Faster-CNN object detection model. Our experimentation result comparison consisting of three main phases. In phase 1, we provided the result comparison of state of art model along with YOLO-V3 model without utilizing transfer learning technique. In phase 2, we provided the result comparison which explores the data augmentation technique infused state of art model along with data augmentation technique infused YOLO-V3 without utilizing transfer learning. In phase 3, we provided the result comparison which explores the data augmentation technique infused state of art model along with data augmentation technique infused YOLO-V3 with utilizing transfer learning.

Table 2 clearly indicates the working performance of state of art model and YOLO-V3 object detection methods without utilizing transfer learning technique. As the result, we can see the working progress all YOLO-V3 model along with state of art model in terms of various performance evaluation metrices. From the comparison YOLO-V3 model has achieved with highest rate of accuracy. Respective comparison graph is given in Fig. 2.

Once data augmentation is applied, then we compared proposed YOLO-V3 model against other state of art model. Our proposed YOLO-ctV3 with transfer learning technique achieves the highest accuracy of 94.09 % while classify tumor as stage 1, stage 2 and stage 3. And the respective graph as given in Fig. 3. The metrics that we are used to compare is accuracy, precision, recall and f-measure. As the result proves that our proposed technique achieves highest performance in terms of all metrics that we are used. The respective comparison Table 3 and graph are given as follow,

Finally, we compared proposed YOLO-V3 model (with transfer learning) which is infused of data augmented technique with other state of art models. Here transfer learning techniques adopt two fully connected layer in which we trained those layers using CVC colon dataset, which is mentioned in description of dataset, which is mentioned in description of dataset. Our proposed YOLO-V3 with transfer learning

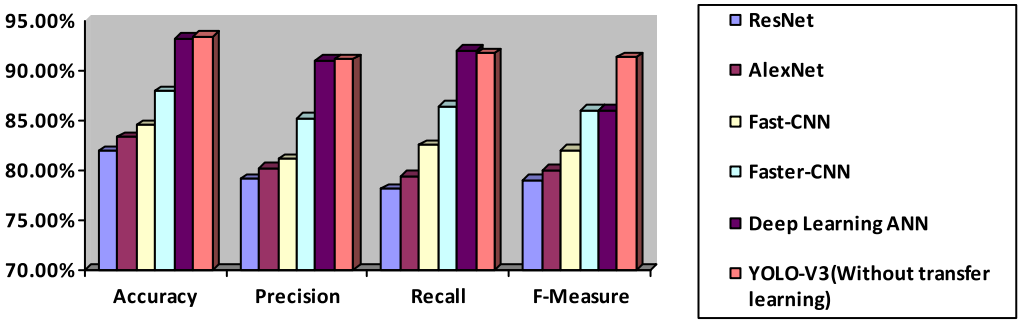


Fig. 2. Comparison between YOLO-V3(Without transfer learning) against other state of art models.

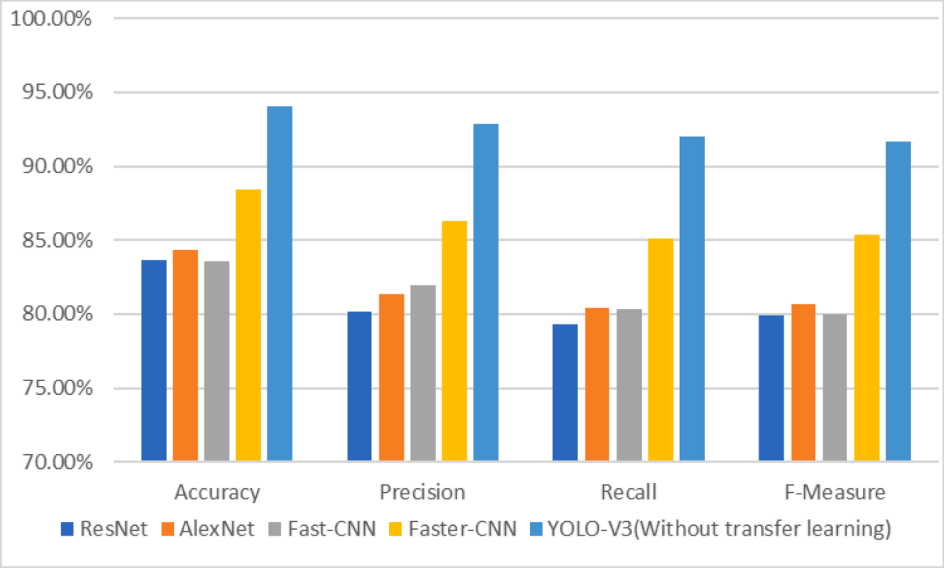


Fig. 3. Comparison between data augmentation technique infused YOLO-V3 (without transfer learning) model against state of art models.

Table 3
Comparison between data augmentation technique infused YOLO-V3(Without transfer learning) model against state of art models.

Model	Accuracy	Precision	Recall	F-Measure
ResNet	83.63 %	80.13 %	79.33 %	79.88 %
AlexNet	84.33 %	81.34 %	80.39 %	80.64 %
Fast-CNN	83.53 %	81.98 %	80.35 %	80.03 %
Faster-CNN	88.45 %	86.33 %	85.12 %	85.32 %
YOLO-V3(Without transfer learning)	94.09 %	92.84 %	91.98 %	91.64 %

technique achieves the highest accuracy of 96.04 % while classify tumor as stage 1, stage 2 and stage 3. And the respective graph as given in Fig. 4. The metrics that we are used to compare is accuracy, precision, recall and f-measure. As the result proves that our proposed technique achieves highest performance in terms of all metrics that we are used. The respective comparison Table 4 and graph are given as follow,

5. Conclusion

Colo-Rectal Cancer (CRC) stood witnessed as the major cause of deaths worldwide, especially in men. But the early detection and removal of polyps can reduce the death rate to a considerable extent. Hence present study, we introduced YOLOv3 Multi-Scale Framework

Table 4
Comparison between data augmentation technique infused YOLO-V3(with transfer learning) model against state of art models.

Model	Accuracy	Precision	Recall	F-Measure
ResNet	83.63 %	80.13 %	79.33 %	79.88 %
AlexNet	84.33 %	81.34 %	80.39 %	80.64 %
Fast-CNN	83.53 %	81.98 %	80.35 %	80.03 %
Faster-CNN	88.45 %	86.33 %	85.12 %	85.32 %
YOLO-V3(With transfer learning)	96.04 %	93.98 %	94.87 %	94.02 %

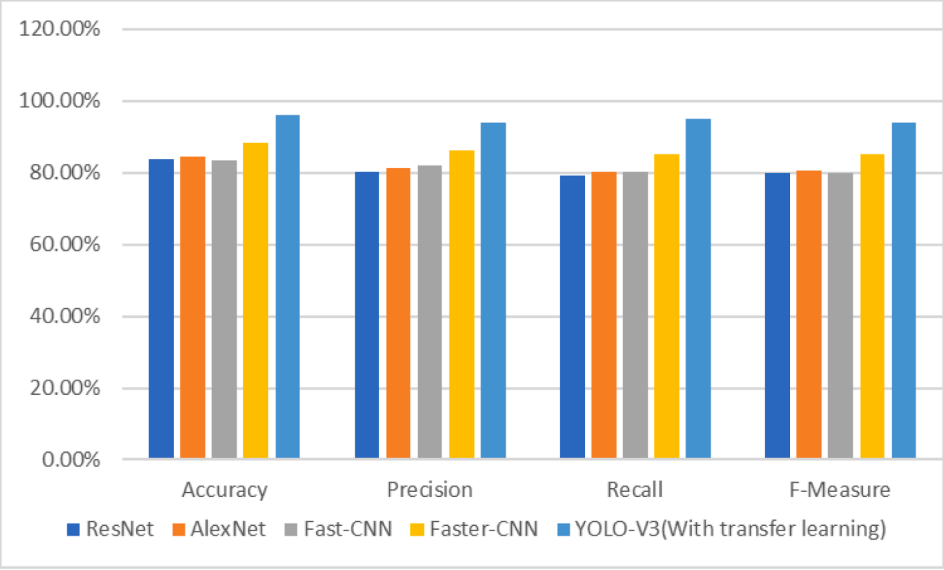


Fig. 4. Comparison between data augmentation technique infused YOLO-V3(with transfer learning) model against state of art models.

(YOLOv3-MSF) in order to detect tumor, once tumor get detected then it is classified as stage 1, stage 2 or stage 3 based on its length. Present study utilizes CVC colon dataset, and also transfer learning technique being adopted in order to train the dataset. We used two fully connected layer in order to provide additional training over the existing architecture. In order to evaluate the performance of the proposed system, various metrics like accuracy, precision, recall and f-measures are being used. Also, comparison has to be done, with state of art model like Faster-CNN. As the result proves that our proposed YOLOv3 with transfer learning technique performed well in terms of different evaluation metrics. Also, we aim to integrate layers like LSTM and Regression layer in the existing architecture in our future work.

CRediT authorship contribution statement

Malathi Murugesan: Conceptualization, Methodology, Software, Data curation, Writing – original draft, Visualization, Investigation, Software, Validation. **R. Madonna Arieth:** . **Shankarlal Balraj:** . **R. Nirmala:** .

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References:

[1] Stewart BW, Wild CP: World cancer report 2014. http://publications.iarc.fr/Non-SeriesPublications/World-Cancer-Reports/WorldCancer-Report-2014 (accessed February 10, 2018).

[2] S.J. Winawer, A.G. Zauber, M.N. Ho, M.J. O'Brien, L.S. Gottlieb, S.S. Sternberg, et al., Prevention of colorectal cancer by colonoscopic polypectomy. The National Polyp Study Workgroup, N. Engl. J. Med. 329 (1993) 1977–1981.

[3] O. Kronborg, Colon polyps and cancer, Endoscopy 36 (2004) 3–7.

[4] K. Gono, M. Igarashi, T. Obi, M. Yamaguchi, N. Ohyama, Multiple-discriminant analysis for light-scattering spectroscopy and imaging of two-layered tissue phantoms, Opt Lett 29 (2004) 971.

[5] K. Gono, T. Obi, M. Yamaguchi, N. Ohyama, H. Machida, Y. Sano, et al., Appearance of enhanced tissue features in narrow-band endoscopic imaging, J Biomed Opt 9 (2004) 568.

[6] H. Machida, Y. Sano, Y. Hamamoto, M. Muto, T. Kozu, H. Tajiri, et al., Narrow-band imaging in the diagnosis of colorectal mucosal lesions: a pilot study, Endoscopy 36 (2004) 1094–1098.

[7] S. Tanaka, K. Haruma, C.R. Teixeira, S. Tatsuta, N. Ohtsu, Y. Hiraga, et al., Endoscopic treatment of submucosal invasive colorectal carcinoma with special reference to risk factors for lymph node metastasis, J Gastroenterol 30 (1995) 710–717.

[8] K. Kitajima, T. Fujimori, S. Fujii, J. Takeda, Y. Ohkura, H. Kawamata, et al., Correlations between lymph node metastasis and depth of submucosal invasion in submucosal invasive colorectal carcinoma: a Japanese collaborative study, J Gastroenterol 39 (2004) 534–543.

[9] T. Watanabe, K. Muro, Y. Ajioka, Y. Hashiguchi, Y. Ito, Y. Saito, et al., Japanese Society for Cancer of the Colon and Rectum (JSCCR) guidelines 2016 for the treatment of colorectal cancer, Int J Clin Oncol (2017), https://doi.org/10.1007/s10147-017-1101-6.

[10] A. Ignjatovic, J.E. East, N. Suzuki, M. Vance, T. Guenther, B.P. Saunders, Optical diagnosis of small colorectal polyps at routine colonoscopy (Detect InSpect ChAracterise Resect and Discard; DISCARD trial): a prospective cohort study, Lancet Oncol 10 (2009) 1171–1178.

[11] L. Deng, D. Yu, Deep learning: methods and applications, FnT Signal Processing 7 (2013) 197–387.

[12] M. Løberg, M. Kalager, Ø. Holme, G. Hoff, H.-O. Adami, M. Bretthauer, Long-term colorectal-cancer mortality after adenoma removal, N Engl J Med 371 (2014) 799–807.

[13] Y. Sano, S. Tanaka, S. Kudo, S. Saito, T. Matsuda, Y. Wada, et al., Narrow-band imaging (NBI) magnifying endoscopic classification of colorectal tumors proposed by the Japan NBI Expert Team, Dig Endosc 28 (2016) 526–533.

[14] M. Iwatate, Y. Sano, S. Hattori, W. Sano, N. Hasuike, T. Ikumoto, et al., The addition of high magnifying endoscopy improves rates of high confidence optical diagnosis of colorectal polyps, Endosc Int Open 3 (2015) E140–E145.

[15] T. Hisabe, S. Tsuda, T. Hoashi, H. Ishihara, K. Yamasaki, T. Yasaka, et al., Validity of conventional endoscopy using "non-extension sign" for optical diagnosis of colorectal deep submucosal invasive cancer, Endosc Int Open 6 (2018) E156–E164.

[16] G. Schachschal, M. Mayr, A. Treszl, K. Balzer, K. Wegscheider, J. Aschenbeck, et al., Endoscopic versus histological characterisation of polyps during screening colonoscopy, Gut 63 (2014) 458–465.

[17] A. Boikov, V. Payor, R. Savelev, A. Kolesnikov, Synthetic data generation for steel defect detection and classification using deep learning, Symmetry 13 (7) (2021) 1176.

[18] Liu L, Ouyang W, Wang X, Fieguth P, Liu X, Pietikäinen M (2018) Deep learning for generic object detection: a survey. arXiv preprint https://arxiv.org/abs/1809.02165v4.

[19] H. Zhang, Y. Ji, W. Huang, et al., Sitcom-star-based clothing retrieval for video advertising: a deep learning framework, Neural Comput Appl 31 (2019) 7361–7380.

[20] M.K. Benkaddour, S. Lahlali, M. Trabelsi, Human Age and Gender Classification using Convolutional Neural Network, in: In 2020 2nd International Workshop on Human-Centric Smart Environments for Health and Well-beIng (IHSH), IEEE, 2021, pp. 215–220.

[21] Attia A, Dayan S (2018) Detecting and counting tiny faces. arXiv preprint https://arxiv.org/abs/1801.06504.

[22] L. Schoneveld, A. Othmani, H. Abdelkawy, Leveraging recent advances in deep learning for audio-visual emotion recognition, Pattern Recogn. Lett. 146 (2021) 1–7.

[23] E. Ramanujam, T. Perumal, S. Padmavathi, Human activity recognition with smartphone and wearable sensors using deep learning techniques: A review, IEEE Sens. J. 21 (12) (2021) 13029–13040.

[24] Zhou Y, Liu L, Shao L, Mellor M (2016) DAVE: a unified framework for fast vehicle detection and annotation. arXiv preprint https://arxiv.org/abs/1607.04564.

[25] Wang L, Lu Y, Wang H, Zheng Y, Ye H, Xue X (2017) evolving boxes for fast vehicle detection. arXiv preprint https://arxiv.org/ abs/1702.00254.

[26] S.M. Azimi, R. Bahmanyar, C. Henry, F. Kurz, Eagle: Large-scale vehicle detection dataset in real-world scenarios using aerial imagery, in: In 2020 25th International Conference on Pattern Recognition (ICPR), IEEE, 2021, pp. 6920–6927.

[27] Chen X, Ma H, Wan J, Li B, Xia T (2016) Multi-view 3D object detection network for autonomous driving. arXiv preprint https:// arxiv.org/abs/1611.07759.

[28] Uc¸ar A, Demir Y, Gu¨zeli C (2017) Object recognition and detection with deep learning for autonomous driving applications. Simulation 93:759–769.

[29] Nguyen-Meidine LT, Granger E, Kiran M, Blais-Morin LA (2018) A comparison of cnn-based face and head detectors for real-time video surveillance applications. arXiv preprint https:// arxiv.org/abs/1809.03336.

[30] Yu R, Wang H, Davis LS (2018) ReMotENet: efficient relevant motion event detection for large-scale home surveillance videos. arXiv preprint https://arxiv.org/abs/1801.02031.

[31] J. Shen, N. Liu, H. Sun, Vehicle detection in aerial images based on lightweight deep convolutional network, IET Image Proc. 15 (2) (2021) 479–491.

[32] Redmon J, Farhadi A (2018) YOLOv3: An incremental improvement. In: Proceedings of the IEEE conference on computer vision and pattern recognition. arXiv preprint https://arxiv. org/abs/1804.02767.

[33] W. Liu, D. Anguelov, D. Erhan, et al., SSD: single shot multi box detector, in: In: Proceedings of the European conference on computer vision, 2016, pp. 21–37.

[34] C.Y. Fu, W. Liu, A. Ranga, et al., DSSD: deconvolutional single shot detector. In: Proceedings of the IEEE conference on computer vision and pattern recognition. arXiv preprint https:// arxiv.org/abs/1701.06659, 2017.

[35] T. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, S. Belongie, Feature pyramid networks for object detection, in: In: Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 2117–2125.

[36] Shi W, Caballero J, Huszár F, Totz J, Aitken AP, Bishop R et al (2016) Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In: Proceedings of the IEEE conference on computer vision and pattern recognition.

[37] Lakshminarayanan, K., N. Muthukumaran, Y. Harold Robinson, Vimal Shanmuganathan, Seifedine Kadry, and Yunyoung Nam. "Deep learning-based hookworm detection in wireless capsule endoscopic image using AdaBoost classifier." (2021).

[38] Ngoc Lan, Phan, Nguyen Sy An, Dao Viet Hang, Dao Van Long, Tran Quang Trung, Nguyen Thi Thuy, and Dinh Viet Sang. "NeoUNet: Towards accurate colon polyp segmentation and neoplasm detection." In International Symposium on Visual Computing, pp. 15-28. Springer, Cham, 2021.

[39] B. V. Dhandra, R. Hegadi, M. Hangarge, and V. S. Malemath, "Analysis of abnormality in endoscopic images using combined hsi color space and watershed segmentation," in Proc. 18th Int. Conf. Pattern Recognit. (ICPR), vol. 4, Aug. 2006, pp. 695–698.

[40] S. Hwang and M. E. Celebi, "Polyp detection in wireless capsule endoscopy videos based on image segmentation and geometric feature," in Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP), Mar. 2010, pp. 678–681.

[41] Y. Wang, W. Tavanapong, J. Wong, J. Oh, P.C. de Groen, 'Partbased multiderivative edge cross-sectional profiles for polyp detection in colonoscopy', IEEE J. Biomed. Health Informat. 18 (4) (Jul. 2014) 1379–1389.

[42] N. Tajbakhsh, S. R. Gurudu, and J. Liang, "Automatic polyp detection in colonoscopy videos using an ensemble of convolutional neural networks," in Proc. IEEE 12th Int. Symp. Biomed. Imag. (ISBI), Apr. 2015, pp. 79–83.

[43] N. Tajbakhsh, J.Y. Shin, S.R. Gurudu, R.T. Hurst, C.B. Kendall, M.B. Gotway, J. Liang, 'Convolutional neural networks for medical image analysis: Full training or fine tuning?', IEEE Trans. Med. Imag. 35 (5) (May 2016) 1299–1312.

[44] R. Zhang, Y. Zheng, T.W.C. Mak, R. Yu, S.H. Wong, J.Y.W. Lau, C.C.Y. Poon, 'Automatic detection and classification of colorectal polyps by transferring low-level CNN features from nonmedical domain', IEEE J. Biomed. Health Inform. 21 (1) (Jan. 2017) 41–47.

[45] J. Bernal, et al., 'comparative validation of polyp detection methods in video colonoscopy: Results from the miccai 2015 endoscopic vision challenge', IEEE Trans. Med. Imag. 36 (6) (Jun. 2017) 1231–1249.

[46] P. Wang, S. M. Krishnan, C. Kugean, and M. P. Tjoa, "Classification of endoscopic images based on texture and neural network,'' in Proc. Conf. 23rd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc., vol. 4, Oct. 2001, pp. 3691–3695.
[47] P. Li, K. L. Chan, and S. M. Krishnan, "Learning a multi-size patchbased hybrid kernel machine ensemble for abnormal region detection in colonoscopic images,'' in Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR), vol. 2, Jun. 2005, pp. 670–675.
[48] B.E. Bejnordi, et al., 'Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer', JAMA 318 (22) (Dec. 2017) 2199–2210.
[49] V. Gulshan, L. Peng, M. Coram, M. Stumpe, D. Wu, A. Narayanaswamy, S. Venugopalan, K. Widner, T. Madams, J. Cuadros, R. Kim, R. Raman, P. Nelson, J. Mega, D. Webster, 'Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs', J. Amer. Med. Assoc. 316 (22) (2016) 2402–2410.
[50] L. Yu, H. Chen, Q. Dou, J. Qin, P.A. Heng, 'Integrating online and offline three-dimensional deep learning for automated polyp detection in colonoscopy videos', IEEE J. Biomed. health Informat. 21 (1) (Jan. 2017) 65–75.
[51] H. Chen, X. J. Qi, J. Z. Cheng, and P. A. Heng, "Deep contextual networks for neuronal structure segmentation,'' in Proc. AAAI, 2016, pp. 1167–1173.
[52] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A.C. Berg, L. Fei-Fei, 'ImageNet large scale visual recognition challenge', Int. J. Comput. Vis. 115 (3) (Dec. 2015) 211–252.
[53] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition,'' in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2016, pp. 770–778.
[54] E. Shelhamer, J. Long, T. Darrell, 'Fully convolutional networks for semantic segmentation', IEEE Trans. Pattern Anal. Mach. Intell. 39 (4) (Apr. 2017) 640–651.
[55] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single shot multibox detector,'' in Proc. 14th Eur. Conf. Comput. Vis. (ECCV). Amsterdam, The Netherlands: Springer, Oct. 2016, pp. 21–37.
[56] A. Juneja, S. Juneja, A. Soneja, S. Jain, Real time object detection using CNN based single shot detector model, J. Inform. Technol. Manage. 13 (1) (2021) 62–80.
[57] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection,'' Jun. 2016, pp. 779–788.