Retraction

Retraction: Yolov3 Supervised Machine Learning Framework for Real-Time Object Detection and Localization (J. Phys.: Conf. Ser. 1916 012032)

Published 23 February 2022

This article (and all articles in the proceedings volume relating to the same conference) has been retracted by IOP Publishing following an extensive investigation in line with the COPE guidelines. This investigation has uncovered evidence of systematic manipulation of the publication process and considerable citation manipulation.

IOP Publishing respectfully requests that readers consider all work within this volume potentially unreliable, as the volume has not been through a credible peer review process.

IOP Publishing regrets that our usual quality checks did not identify these issues before publication, and have since put additional measures in place to try to prevent these issues from reoccurring. IOP Publishing wishes to credit anonymous whistleblowers and the Problematic Paper Screener [1] for bringing some of the above issues to our attention, prompting us to investigate further.

[1] Cabanac G, Labbé C and Magazinov A 2021 arXiv:2107.06751v1

Retraction published: 23 February 2022
Yolov3 Supervised Machine Learning Framework for Real-Time Object Detection and Localization

Srithar S1, Priyadharsini M2, Margret Sharmila F3, Rajan R4
1Assistant Professor, Department of CSE, KPR Institute of Engineering and Technology, Coimbatore
2Assistant Professor, Department of IT, Sri Ramakrishna Engineering College, Coimbatore
3Assistant Professor, Department of CSE, SNS College of Engineering, Coimbatore
4Assistant Professor, Department of CSE, KGiSL Institute of Technology, Coimbatore
1sss.srithar@gmail.com, 2priya.murugesan@srec.ac.in,
3margretsharmilait@gmail.com, 4rajanramalingam1986@gmail.com

Abstract. Nowadays to detect and classify the objects from the sequence of image frames various machine learning models are used. The performance of the object recognition model is purely depending on the number of trained images. The image processing and pattern extraction focus on object recognition, localization, and classification. To classify the objects from the external events or to detect multiply objects from an image the confident level of the trained weights should be maximum. The existing object detection methods look at the specific region to detect and classify the objects in an image. The proposed object detection system takes an improved YOLOv3 model for object classification. The improved model will look at the entire image to detect and recognize the objects. The YOLOv3 model uses the neural network on the image which split the image into a region and map the confidence probability. The proposed model will detect multiple objects by exploiting the contextual information using a single CNN. The model can process 45 frames in a second and it is suitable for object detection in real-time.

Keywords: bounding box, object detection, YOLO model, grid, classification, localization

1. Introduction
The application of computer vision integrated machine learning is mostly used in artificial intelligence irrespective of the engineering field. Some of the applications are gender classification, vehicle detection, pose detection, mask detection, and gun detection, etc. The framework convolutional neural network is mostly used to detect and locate the objects from the image or video. The object detection algorithm will formulate the bounding box to locate the object whereas the classification algorithms will differentiate the objects like car, cell phone, bottle, person and chair, etc [1-4]. The CNN model is expanded into region proposed CNN to improve the prediction accuracy. The expanded neural network methods like R-CNN, fast CNN, and Faster CNN are used for object detection. The standard CNN may produce less accurate results due to the length of the neural connection and region of object interest. A challenging issue in object detection is the pose of the trained image and the testing image may be variant.
2. Related Work

[5] investigated R-CNN to detect the objects in an image. The R-CNN prediction model uses the selective search approach to catalog 2000 regions from an image. The input image is extracted into 2000 region proposals and each region is compared with the trained objects. The neural network will map all the regions into the trained model and classify the region by the match probability. But the algorithm will take much time to train the system because the algorithm needs to execute 2000 regions per image. The algorithm will take approximately 40-50 seconds to classify the region of interest. The selective search algorithm is a defined version where we cannot able to import the reinforcement learning approaches.

[6] investigated fast R-CNN to improve the training accuracy. The fast R-CNN model will not take the region proposals into the CNN for feature extraction. Instead, the entire image segment is given into the CNN to generate the feature map, so that the convolution takes only one time per image. It also uses the selective search algorithm to detect the region of interest.

[7] investigate the Faster R-CNN model by executing four steps in a row. The first step is to obtain the region proposal to detect and classify the object bounding box. Second, the algorithm will extract the most relevant features from the images. Third, the classification algorithm identifies the category of the bounded object. Finally, the regression layer draws a closed line among the bounding box.

[8] investigated object recognition by considering the image as several fragments. In this approach, the system will not take the image as a whole to detect the object. The CNN model will compare each segment with the trained model which is a very difficult process and the time complexity for processing is more [9-14].

3. YOLOv3 Supervised Machine Learning Framework

The YOLOv3 object detection process flow contains 5 different steps. The process flow of the YOLOv3 object detection model is represented in Figure 1.

![YOLOv3 object detection model process flow](image)

**Figure 1.** YOLOv3 object detection model process flow

3.1 Image Capture

The object detection model starts with image capture from the real-time video. Mostly in a second 24-60 frames, distinct images are represented.

3.2 Grid Formation (S*S size)
Once the image is captured, it is divided by \((S*S)\) grid size. The classification and localization are applied on each grid cell to detect the objects. The size of the image is represented as 100\%*100\%. Figure 2 shows the 3*3 grid formulation for the image frame.

![Figure 2. S*S grid formulation from an image](image)

### 3.3 Grid Bounding Box

The YOLO algorithm will detect the bounding boxes from each grid cell. Each grid cell determines the value parameter in terms of eight dimensions namely object probability, bounding box height, width, and initial point, finally classes. The value \(C_p\) represents the confidence probability of the objects. The values \(bX\), \(bY\), \(bH\), and \(bW\) are represented as the bounding box initial position, height, and width. There are 92 different classes are defined in YOLO represented from \(c_1\), \(c_2\), … \(c_n\). Figure 3 shows the template for the grid structure, Figure 4 shows the \(A_{21}\) frame of the given image frame, Figure 5 shows the mathematical feature of the selected frame.

![Figure 3. A grid mathematical feature](image)

![Figure 4. A_{21} frame for the sample](image)
3.4 Bounding Box Combination

The actual bounding box and the predicted bounding box over the image and their difference are used to predict the confidence rate of the bounding box. If the difference is greater than 0.5 then the prediction is good.

3.5 Confidence Estimation

The proposed model detects several bounding boxes in an image and confidence probability. The model takes the entire image and predicts the coordinates of the bounding box, confidence rate. The YOLO model will train the image by the whole. The model can effectively process 45fps and the faster version can take up to 150fps. The existing fast R-CNN model may detect the background image due to the large context. The proposed object detection will eliminate the error rate up to 50 percent. The convolution network will detect the features to perceive the bounding box. Each input image split the image into a segment of the ZxZ grid which is responsible to detect the objects. Also, the grid cell will rightly point out the bounding box and prediction accuracy score. If no object is detected within the bounding box then the confidence score is assumed as zero.

\[ X = \Pr(\text{object}) \times \text{IOU} \]

The bounding box contains 5 primary components. The bounding box origin is represented by \((x,y)\), width\((w)\), height\((h)\) and the confident rate\((c)\). The confident score is calculated for each of the grid cells.

3.6 Final Decision (detected object)

Finally, the classes are represented based on the confidence rate. The confidence threshold is assumed as 0.7 if multiple objects are detected within a frame. The possible trained objects are listed and shown in Figure 6.

Person, bottle, cup, bicycle, donut, spoon, skateboard, book, bus, bowl, motorbike, apple, car, truck, traffic light, toothbrush, kite, baseball bat, chair, tennis racket, broccoli, stop sign, sandwich, aeroplane, banana, parking meter, hot dog, baseball glove, carrot, train, toaster, surfboard, refrigerator, umbrella, scissors, cat, pizza, bench, sofa, sheep, wine glass, bed, dining table, backpack, orange, horse, fire hydrant, cake, hair drier, cow, tv monitor, mouse, sports ball, knife, toilet, teddy bear, snowboard, laptop, boat, potted plant, microwave oven, sink, bear, handbag, suitcase, skis, keyboard, elephant, dog, clock, tie, bird, vase, giraffe, frisbee, cell phone, zebra, fork, remote

Figure 6. YOLO trained objects
4. Results and Discussion

The experiment is simulated in the Anaconda3 python framework. The system will take yolov3.weights and yolov3.cfg to detect the objects. The defined YOLO model is trained with 92 different objects represented in Figure 7. The YOLOv3 model predicts the object such as a person, cell phone, and bottle from the real-time video. The proposed approach is incorporated into three different applications namely real-time object detection, age and gender classification.

The YOLOv3 model also detects the human face and age category by extracting the features. The system also detects multiple faces on the video frames by enabling different bounding boxes. The bounding box may intersect other boxes. The system is trained with different age groups such as (0-2), (4-6), (8-12), (15-20), (25-32), (38-43), (48-53), and (60-100). The confidence level of the recognized objects should be greater than 0.7. If more than one match is found, the confidence probability is taken for the image which is the maximum confidence rate. The YOLO gender and age classification is represented in Figure 8. The confident and object classification has experimented in Figure 9. A random object classification, average confidence, and detection time are represented in Table 1.
Figure 9. Age and Gender classification confidence rate for image frames

Table 1. Sample classification rate

| S.No. | Classification Gender | Age  | Average Confidence | Time (in milli seconds) |
|-------|-----------------------|------|--------------------|------------------------|
| 1     | Male                  | [0-2]| 0.933              | 0.067                  |
| 2     | Male                  | [0-2]| 0.945              | 0.078                  |
| 3     | Male                  | [25-32]| 0.967             | 0.086                  |
| 4     | Female                | [25-32]| 0.912             | 0.063                  |
| 5     | Male                  | [25-32]| 0.930             | 0.040                  |
| 6     | Male                  | [25-32]| 0.989             | 0.057                  |
| 7     | Male                  | [0-2]| 0.904              | 0.089                  |
| 8     | Male                  | [0-2]| 0.912              | 0.054                  |
| 9     | Male                  | [0-2]| 0.867              | 0.683                  |
| 10    | Male                  | [0-2]| 0.867              | 0.602                  |
| 11    | Female                | [25-32]| 0.987             | 0.678                  |
| 12    | Female                | [25-32]| 0.865             | 0.734                  |
| 13    | Male                  | [25-32]| 0.923             | 0.756                  |
| 14    | Male                  | [0-2]| 0.789              | 0.790                  |
| 15    | Male                  | [0-2]| 0.756              | 0.867                  |
| 16    | Female                | [25-32]| 0.878             | 0.780                  |
| 17    | Male                  | [0-2]| 0.965              | 0.935                  |
| 18    | Male                  | [25-32]| 0.896             | 0.786                  |

5. Conclusion

The application of computer vision and supervised learning models is widely used in artificial intelligence. To detect and locate the object from the real-time video or image is more challenging. To improve the prediction accuracy in terms of classification and localization the model YOLOv3 is formulated. In the image segment, the object is located within the bounding boxes along with the confidence rate. In our approach the whole image is taken to detect and recognize the features. The proposed method improved the prediction accuracy and the confidence rate up to 94.34 percent.
References:

[1] Ahmad, A., Jaka, AP., Arlan, AG., Detection and classification of vehicles for traffic video analytics, *Procedia Computer Science*, Vol.144, pp. 259-268, 2018.

[2] Can, P., Kun, Z., Brian CL., Faster ILOD: Incremental learning for object detectors based on faster RCNN, *Pattern Recognition Letters*, Vol.140, pp. 109-115, 2020.

[3] Chunlong, X., Ping, W., Wenwen, W., Nanning, Z., A multilevel fusion network for 3D object detection, *Neurocomputing*, Vol. 437, pp. 107-117, 2021.

[4] Fulong, L., Gang, L., Shuqiang, Y., Wenjuan, Y., Guoquan, H., Ling, L., Detection of heterogeneity in multi-spectral transmission image based on spatial pyramid matching model and deep learning, *Optics and Lasers in Engineering*, Vol. 134, 106272, 2020.

[5] Hanyu, P., Shifeng, C., BDNN: Binary convolution neural networks for fast object detection, *Pattern Recognition Letters*, Vol. 125, pp. 91-97, 2019.

[6] Lin, J., Shengyu, Z., Shifeng, D., Hongqiang, W., RFP-Net: Receptive field-based proposal generation network for object detection, *Neurocomputing*, Vol. 405, pp. 138-148, 2020.

[7] Lin, Z., He, Peilin, Zhang, Changhua, Zhang, Zhehan, A robust approach to reading recognition of pointer meters based on improved mask-RCNN, *Neurocomputing*, Vol. 388, pp. 90-101, 2020.

[8] S., D., & H., A. (2019). AODV Route Discovery and Route Maintenance in MANETs. 2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS). doi:10.1109/icaccs.2019.8728456

[9] H. Anandakumar and K. Uمامaheswari, An Efficient Optimized Handover in Cognitive Radio Networks using Cooperative Spectrum Sensing, *Intelligent Automation & Soft Computing*, pp. 1–8, Sep. 2017. doi:10.1080/10798587.2017.1364931

[10] Rachna, P., Anil, S., Arnav, W., Himanshu, S., Dhiraj, S. An object detection approach for detecting damages in heritage sites using 3-D point clouds and 2-D visual data, *Journal of Cultural Heritage*, 2021.

[11] Roman, S., Weimnn, W., Tatiana, G. Weighted boxes fusion: Ensembling boxes from different object detection models, *Image and Vision Computing*, Vol. 107,104117, 2021.

[12] Tim, H., Christian, S., Moritz, E., Herbert, P., BubCNN: Bubble detection using Faster RCNN and shape regression network, *Chemical Engineering Science*, Vol. 216, 115467, 2020.

[13] Xudong, S., Pengcheng, W., Steven CH., Face detection using deep learning: An improved faster RCNN approach, *Neurocomputing*, Vol. 299, pp. 42-50, 2018.

[14] Zhenbo, L., Ye, L., Yongbo, Y., Ruohao, G., Jiaqi, Y., Jun, Y., Yizhe, W., A high-precision detection method of hydroponic lettuce seedlings' status based on improved Faster RCNN, *Computers and Electronics in Agriculture*, Vol. 182,106054, 2021.