


RESEARCH

Open Access



OSDDY: embedded system-based object surveillance detection system with small drone using deep YOLO

Kaliappan Madasamy¹, Vimal Shanmuganathan¹, Vijayalakshmi Kandasamy¹, Mi Young Lee^{2*}  and Manikandan Thangadurai¹

* Correspondence: miylee@sejong.ac.kr

²Department of Software, Sejong University, Seoul, South Korea
Full list of author information is available at the end of the article

Abstract

Computer vision is an interdisciplinary domain for object detection. Object detection relay is a vital part in assisting surveillance, vehicle detection and pose estimation. In this work, we proposed a novel deep you only look once (deep YOLO V3) approach to detect the multi-object. This approach looks at the entire frame during the training and test phase. It followed a regression-based technique that used a probabilistic model to locate objects. In this, we construct 106 convolution layers followed by 2 fully connected layers and $812 \times 812 \times 3$ input size to detect the drones with small size. We pre-train the convolution layers for classification at half the resolution and then double the resolution for detection. The number of filters of each layer will be set to 16. The number of filters of the last scale layer is more than 16 to improve the small object detection. This construction uses up-sampling techniques to improve undesired spectral images into the existing signal and rescaling the features in specific locations. It clearly reveals that the up-sampling detects small objects. It actually improves the sampling rate. This YOLO architecture is preferred because it considers less memory resource and computation cost rather than more number of filters. The proposed system is designed and trained to perform a single type of class called drone and the object detection and tracking is performed with the embedded system-based deep YOLO. The proposed YOLO approach predicts the multiple bounding boxes per grid cell with better accuracy. The proposed model has been trained with a large number of small drones with different conditions like open field, and marine environment with complex background.

Keywords: Deep learning, Surveillance, YOLO, Convolution neural network, Embedded-based object detection

1 Introduction

The application of drones in various domains is increasing day by day especially military and surveillance to perform deliberate operations in the arena. It is very much important to detect drones in providing security in real-time. Nowadays, real-time detection of drones is a great challenge in various environments like rain, sunlight, and

night. Deep learning plays a major role in detecting objects in various conditions [1]. Recently, computer vision and deep learning approaches such as R-CNN, Faster R-CNN, and mask-RCNN are providing a solution to detect an object [2].

The object detection and tracking system has been variedly applied in various areas of military, health sectors, and security monitoring with autonomy robots [3]. The traditional object identification has been implied with the primary focus on assuming the edges, cuttings, and templates, where the chances of getting the accuracy is lower and the loss is assumed to be of higher [4]. Besides various features extraction methods are been used with CBIR images and various filtering techniques has been applied to the detected objects to ensure the object identification [5]. The gradient-based histogram in image classifiers has been used and the local binary patterns also been applied with the image scanning in the object with sliding window assumptions [6]. The machine learning techniques has been used with the image accuracy enhancement using the PASCAL VOC object detection in handcrafted future. But all of these mechanisms have been challenging phases for the object tracking in the surveillance system using the embedded system [7]. To overcome these challenges, various deep learning models have been proposed to enhance the accuracy [8].

The deep learning models such as R-CNN, Faster R-CNN, and mask-RCNN are not suitable for detecting small objects with speed. In this paper, a novel deep YOLOV3 algorithm is proposed to detect small objects [9]. The proposed work employs confidence score, backbone classifier, and multi-scale predictions to detect small drone objects. The proposed deep YOLOV3 offer the following mechanism for detecting the small drone objects:

The contributions of the paper are as follows:

- This paper proposed a deep YOLOV3 to solve the small object detection issue with speed.
- The confidence score is calculated based on the conditional probability to detect the bounding boxes for a target object.
- Deep YOLOV3 also used a backbone classifier and multi-scale prediction to classify the objects with high accuracy and more accurate surveillance has been achieved.
- The proposed deep YOLOV3 model is achieved 99.99% accuracy to detect the small drone objects with less loss

The rest of the work contains five sections. In Section 2, the paper summarizes the various literature reviews to sort out the problems. In Section 3, the paper is proposed with the deep YOLOV3 model for detecting a small drone object efficiently. In Section 4, the simulation and performance of the proposed deep YOLOV3 model to analyze the loss and accuracy are presented. Section 5 presents the conclusions and future direction.

1.1 Highlights of the proposed YOLOv3 model

- The proposed YOLOv3 uses logistic regression to predict a confidence score for a bounding box and 7×7 grid cells are detected simultaneously. The result shows that this is a very fast model.

- The proposed YOLOv3 uses three anchor boxes to predict three boxes for a grid cell and allocated to a grid cell in detection.
- The proposed YOLOv3 architecture such as 106 convolution layers followed by 2 fully connected layers and $812 \times 812 \times 3$ input size detects a small object with low false alarm and uses minimum filters as possible.
- The proposed YOLOv3 model uses independent logistic classifiers and binary cross-entropy loss during training for small object prediction. It also supports multi-label

2 Related works

Unlu et al. proposed that the commercial unmanned aerial vehicle has been announced as a drone is increasing the modern communication of video and audio with security standards is an essential key factor for the wireless devices [10]. They investigated a novel approach on autonomy drone detection and tracking with a multi-camera angle view [11]. The camera support and the frame have been analyzed with the compact memory and time in an efficient manner [12]. The small-sized aerial intruders are used with the image plane and the compressed images are analyzed with the resource detection algorithm using the deep learning classification algorithms [13, 14].

Deep learning-based object detection has been proposed with various accuracy levels Song Han et al. investigated that the lower cost aerial photography has been used for taking the highlighted pictures and videos with advanced drones and assumed to be of error-prone. A deep drone has been proposed with the embedded system framework where the power drones vision is highly investigated with the drone automatic tracking [15]. The tracking and detection are assumed to be of the multiple hardware with GPU(NVIDIA GTX980) and GPU (NVIDIA Tegra K1 and NVIDIA Tegra X1), the embedded setup has been used to obscure the frame rate, accuracy estimation, and the power assumption has been analyzed with the 1.6 fps for tracking [16]. Redman et al. proposed a new approach on YOLO detectors to use object detection and the work on object classification has been done with the regression problem in bounding boxes and the probability for the association [17]. A neural network has been proposed with the class analysis of probability for the end to end directly in the performance detection and faster R CNN has been used intensively for object detection [18].

Krizhevsky et al. investigated that the deep neural network has been widely used with the computer vision analysis for the binary and multiple image classification and analysis. Besides a tool AlexNet a classic tool of network proposed with image contributes 8 layers and 60 million connections, later VGGNet [8]. Szegedy et al. proposed that GoogleNet has been proposed with the different scaling with the support of CNN. GoogleNet with CNN proposes a convolutional layer with a specified model having 1×1 , 3×3 , and 5×5 with kernel level. The gradient problem has been solved using this multiple cross layers usage [19]. He et al. proposed that ResNet increases the image recognition accuracy and it bypasses the layers that may enhance the absolute values and SqueezeNet has been applied with CNN to enhance the image recognition accuracy with $50\times$ connections and produces with higher accuracy [20].

Henriques et al. highlighted that the kernelized correlation filters have been used for detection of the image classification based on the DFT and a blueprint has been used with the fast algorithms with the resultant in the diagonalized translations in the

storage and computation with the trackers to run 70 frames per second on the NVIDIA TK1 kit [21].

Sabir Hossain et al. highlighted that the target detection and tracking from aerial images have happened with smart sensors and drones. A deep learning-based framework has been proposed with the embedded modules such as Jetson TX or AGX Xavier with Intel Neural computer stick [22]. The flying drones are used with a certain coverage limit and hence it has been followed with the accuracy for estimating the multi-object detection algorithm with the GPU-based embedded for the specified computation power [23]. Deep SORT uses a hypothesis tracking with the support of Kalman filtering with the association metric specified in the multi-rotor drone [24].

Roberto Opromolla highlighted that the UAVs have been used in various civil and military applications specified with the visual cameras that enable the tracking and detection of the track cooperative targets with the frame sequence using deep learning [25]. The you only look once (YOLO), an object detection system, is done with the processing architecture where the machine vision algorithms with the cooperative hints need to be brought to the flight test campaign in the two multirotor UAV. The methods integrate the accuracy and robustness in the challenging environments in the target range vulnerability [26].

Christos Kyrkou et al., proposed a trade-off mechanism with the development of a single-shot object detector using the deep CNN. The UAVs detect the vehicle in a dedicated UAV environment. CNN proposes a holistic approach in the optimization with the deployment UAVs. The aerial images operate in 6–19 frames per second with an accuracy of 95% with the UAV applications with the low power embedded processors [27]. Tsung-Yi Lin et al. highlighted that the RetinaNet performs the object detection using the backbone network and the network supports the classification and regression. The backbone network uses the convolutional feature with the input images combined the Faster R-CNN also uses the Feature Pyramid Network (FPN) [28]. The probability of object presence has been used with the input features mapping in the C channels in the pyramid level with the A anchors and N object classes that take the ReLU activations [29].

Yi Liu et al., highlighted that the UAV has been applied in the tasks of power transmission devices and the usage of deep learning algorithms has been used with the attention of UAV transmission control [30]. The Mask R-CNN has been used with the components of the transmission devices using the edge, hole filling, and hough transform in wireless communication [31]. The accuracy has been achieved with the 100% accuracy in the proposed model using the UAV transmission parameters [32].

LI Y et al. proposed that the multiblock single shot multibox detector (SSD) with the small object detection is used for the surveillance of the railway tracks with UAV. The input images are segmented in terms of patches and the truncated object has been assigned with the sub-layer detection in the two stages with the suppression of sublayer and filtering has been applied in the training samples. The boxes that are not detected in the main layer have been substantially increased with the available SSD and the deep learning model has been proposed with labeling the landslides and the important communication during rainy days has been reported [33].

Jun-Ichiro Watanabe et al. proposed that the YOLO has been applied in the conservation of the marine environments with the micro and macro plastics on land that

occupies the ocean and with that intrusion many species are suspected to face the consequences. The satellite remote sensing techniques for global environmental monitoring have been identified with the object tracker on the ocean surface [34]. Autonomy robots have been used with the observation for the control of the objects in marine environments. The underwater ecosystems have been studied with the learning object algorithm using a DEEP net and the YOLO v3 has been applied and accuracy has been estimated with 77.2% [35]. Kaliappan et al. [36–38] proposed machine learning techniques such as clustering and genetic algorithms to achieve load balancing. Vimal et al. [39] proposed machine learning-based Markov model for energy optimization in cognitive radio networks. Aybora et al. [40] simulated types of annotation errors for object detection using YOLOv3 and examined the erroneous annotations in training and testing phase. Sabir et al. [24] designed a GPU-based embedded flying robot that used a deep learning algorithm to detect and track the real time multiple-object from aerial imagery.

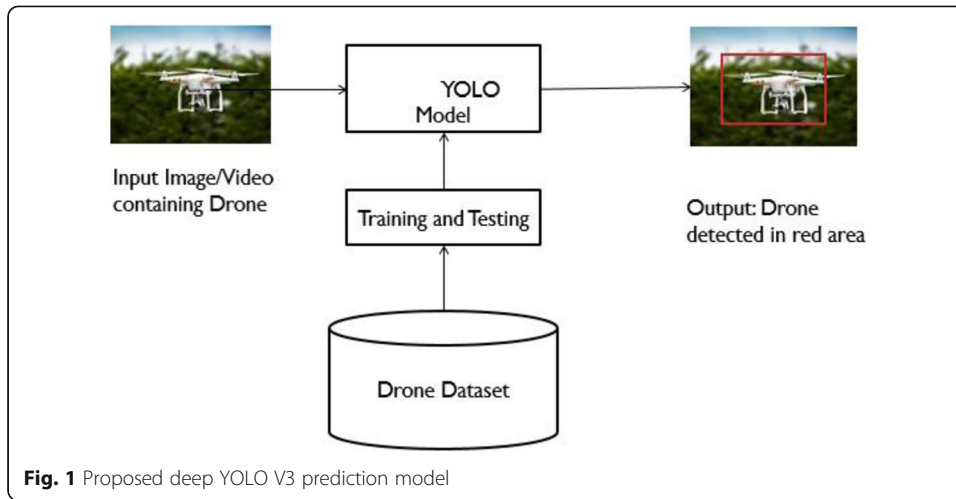
3 Methods

The aim of the proposed model is to analyze the object detection in a real-time environment, with movement decisions using a novel YOLO V3 model using the box within the bounded coordinates. YOLOv3 performs better feature extraction than YOLOv2 because it uses a hybrid approach such as Darknet and residual network. The image is captured and segmented within the bounded box level coordinates. The coordinates are mapped within the box with an interval of frames per second in the novel YOLOV3 model. In this, a deep convolutional neural network (DCNN) to predict with high accuracy has been applied. Various filter sizes such as 32, 64,128,256, and 1024 are applied with striding and padding to process pixel by pixel basis in the frame [2]. The proposed scheme of various sizes kernelized correlation filter (KCF) is used in different convolution layers. In general, KCF runs very fast on video processing. The CNN layers split the image into various regions to predict the accurate bounding box based on the confidence score for all divided regions [41]. The proposed YOLOv3 trained in the Dell EMC workstation which consists of two Intel Xeon Gold 511812 core processor, Six channel memory 256 GB 2666 MHz DDR4 ECC memory, 2× NVIDIA Quadro GV100 GPU and 4× 1TB NVMe class 40 SSD, 1TB SATA HDD.

3.1 Proposed work

The model has been proposed as a novel YOLO V3 deep learning embedded model to detect a small object with a real-time system. YOLO looks at the entire image in an instance to predict the bounding box coordinates [42]. The class probabilities are calculated for all bounding boxes. YOLO can process 45 frames per second. In this, we used a deep convolutional neural network (DCNN) to predict with high accuracy [43]. Figure 1 shows our proposed deep YOLO V3 prediction model to predict the drones. Labeled input images are trained with 45,000 epochs. Each region is a 7×7 grid capable of predicting five bounding boxes. The proposed model detects the drone object as well.

The proposed YOLO algorithm with embedded model used a regression mechanism to predict the classes and bounding boxes for the entire image in every single run in a particular object location. Equation (1) described the bounding box with four



descriptors such as center of a bounding box (b_y, b_h), width (b_w), height (b_h), and a class of an object (c).

$$y = (p_c, b_x, b_y, b_h, b_w, c) \tag{1}$$

The CNN predicts four coordinates for each bounding box such as t_x, t_y, t_w and t_h . The prediction of a bounding box is followed by the following four equations. c_x, c_y is the offset of the top left corner of an image and p_w and p_h are the prior bounding box width and height respectively. Here, t_s is the gradient ground truth value which is computed during the training process.

$$b_x = \sigma(t_x) + c_x \tag{2}$$

$$b_y = \sigma(t_y) + c_y \tag{3}$$

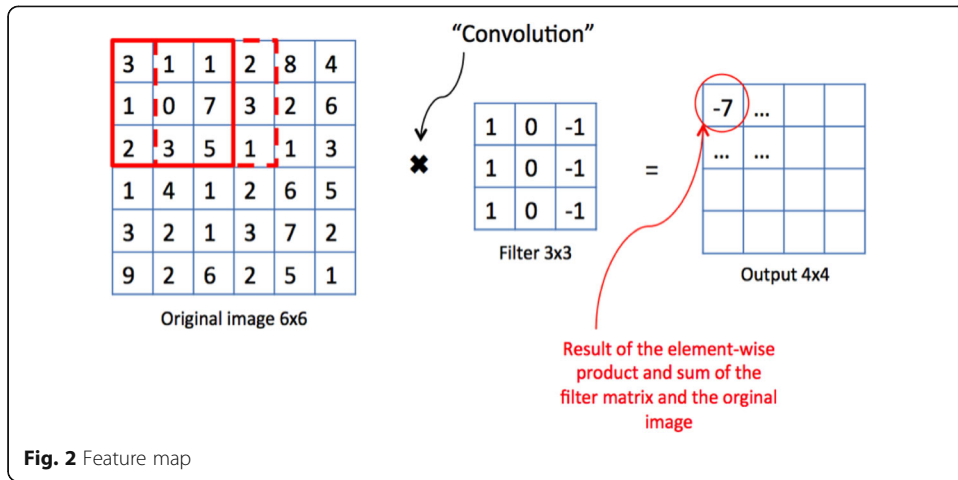
$$b_w = p_w e^{t_w} \tag{4}$$

$$b_h = p_h e^{t_h} \tag{5}$$

The proposed algorithm applied non-max suppression technique called independent logistic regression and predicts the p_c while much of the grid and boxes do not contain a targeted object. p_c is used to predict a confidence score for a bounding box and 7×7 grid cells are detected simultaneously. The result shows that this is a very fast model. This strategy rejects bounding boxes with low probability and predicts a bounding box with the highest probability. The predicted bounding box contains a good confidence score. Equations (6) express the confidence score.

$$p_r(o)x \text{ (IOU)} \tag{6}$$

IOU is an intersection over the union in the region. It is expressed as an area of intersection over the area of the union of two bounding boxes. IOU falls within 0 to 1 and the ground truth box is said to be ~ 1 . IOU is used to find the confidence score for each bounding box that ensures a box contains a predicted target object. IOU also prevents background detection. The confidence score is 0 if there is no object present in the grid cell. Otherwise, the confidence score is equal to IOU between the predicted bounding box and the ground truth box. IOU is achieved greater than 0.5 that ensures



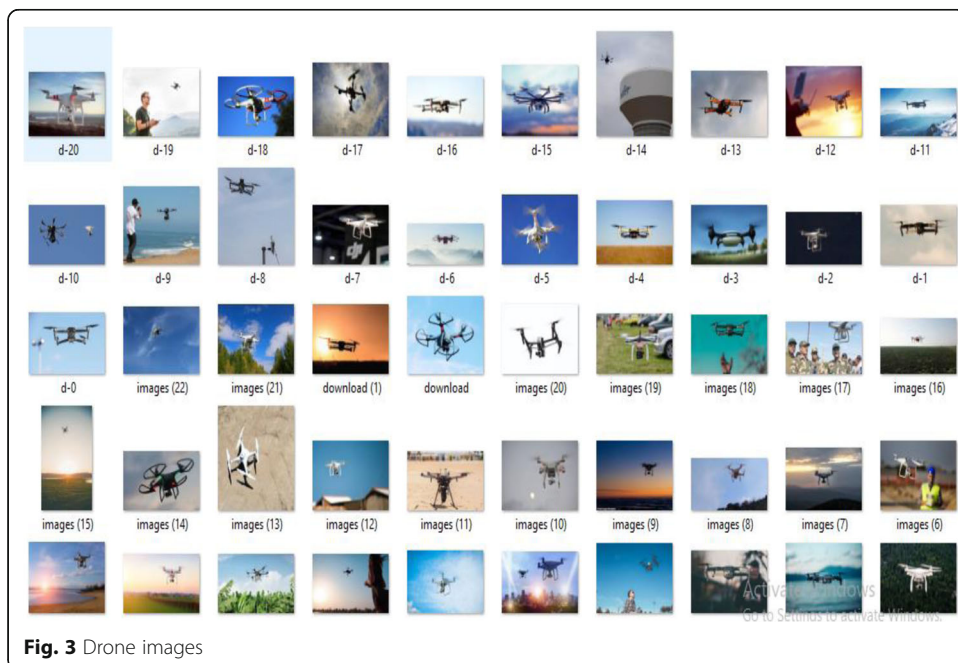
a better prediction with high accuracy for object detection [44]. In order to achieve good prediction, YOLO multiplies the individual box confidence predictions with conditional class probabilities ($p_r(c_i)$) that is expressed in Eq. (7).

$$p_r(c_i | o)p_r(o)IOU_{prediction}^{truth} = p_r(c_i)IOU_{prediction}^{truth} \tag{7}$$

The average IOUs are selected to represent the final predicted bounding boxes which are the much closest prior values for good prediction. It is expressed in the following equation.

$$p_r(o) \times IOU(b, o) = \sigma(t_o) \tag{8}$$

The confidence score of a bounding box plays a vital role in making a prediction at the testing stage. It is the output of a neural network.



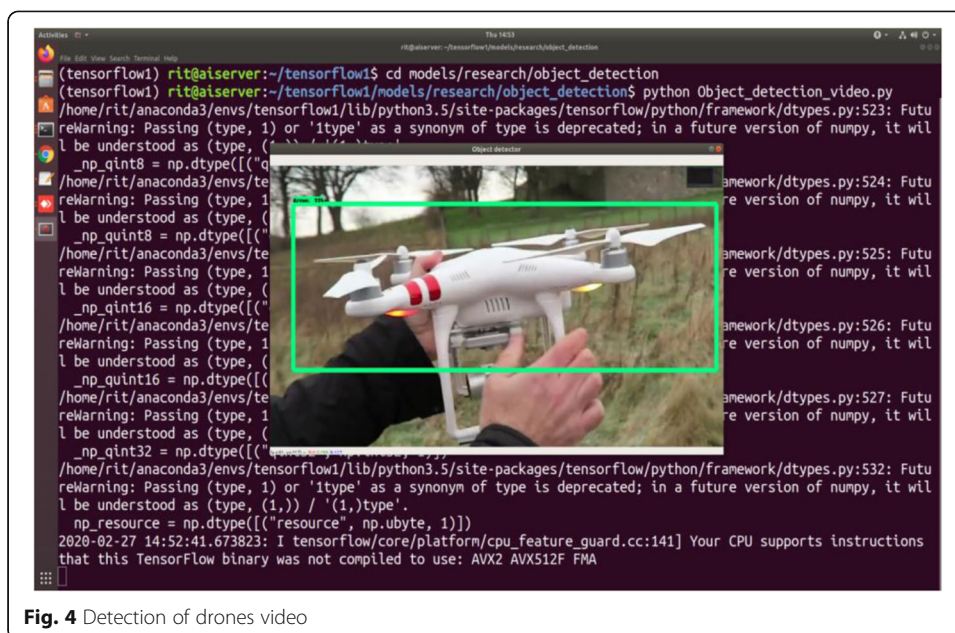


Fig. 4 Detection of drones video

The paper has been designed with a DCNN with 106 convolution layers that contain convolution layers, pooling layers and fully connected layers with classification function. We calculated feature maps with convolution by sliding filter along with the input image in the result is a two-dimension matrix [45]. Figure 2 shows the sample feature map calculated in the convolution layer.

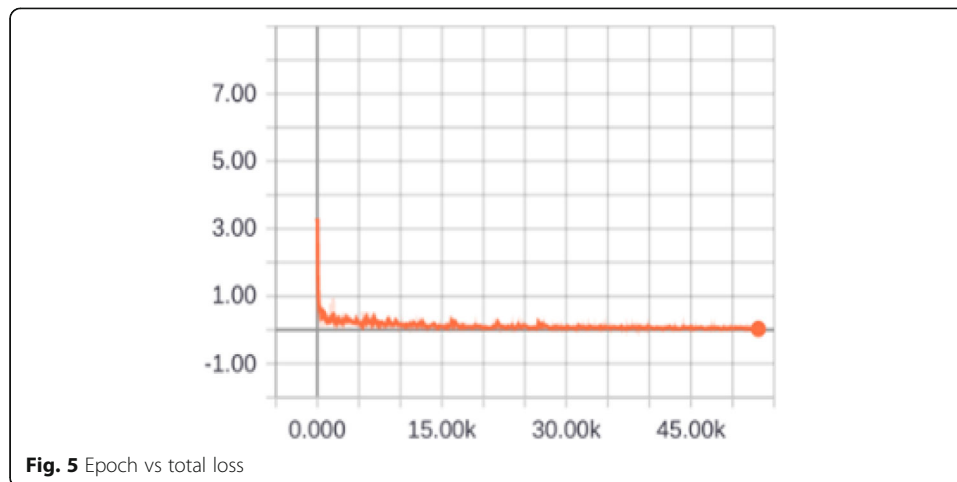
4 Results and discussion

The paper has been proposed with the downloaded 3000 drone images from the Kaggle dataset and Google with around 2 GB. In the proposed YOLOv3, 45,000 epochs have been taken for drone dataset that provide the high accuracy and sensitivity. Figure 3 shows the sample drone images used for training and testing [46]. The proposed work uses a pre-trained YOLOv3 model for training. We implemented the YOLOv3 in GPU-based workstation.

The input images are trained by a pre-trained YOLOv3 model with 106 convolution layers. The training stage takes more than 8 h to build the trained model. This trained YOLO model can accept either image or video input. The proposed model achieved 99.99 percentage results on the detection of drones images or videos [47]. Figure 4 shows the detection of drones' video in the testing stage. Table 1 provides the accuracy comparison of three models such as YOLO, YOLOv2, and YOLOv3. YOLO and YOLOv2 are suitable for large object detection in very fast manner. The proposed

Table 1 Comparison of three YOLO models

Model	Accuracy
YOLO	97.89%
YOLOv2	98.27%
Proposed YOLOv3	99.99%

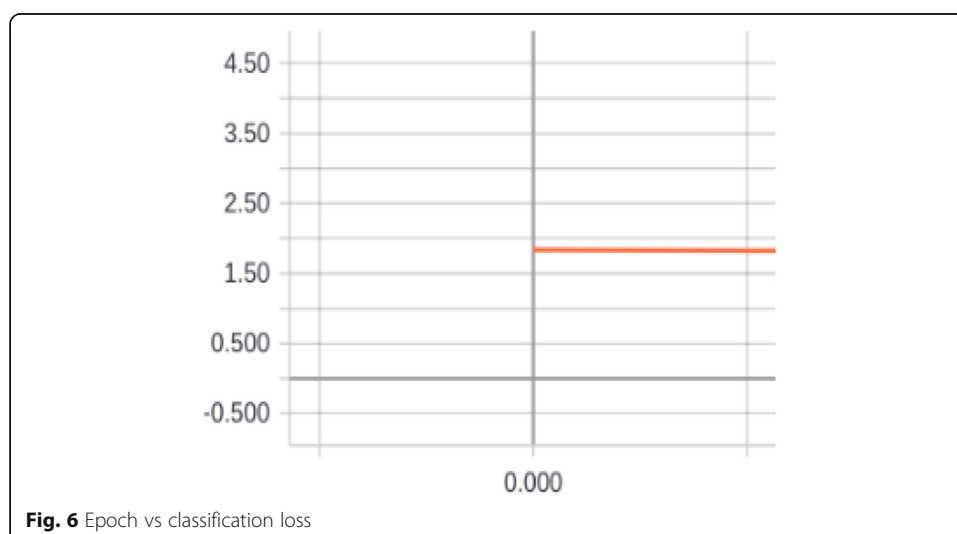


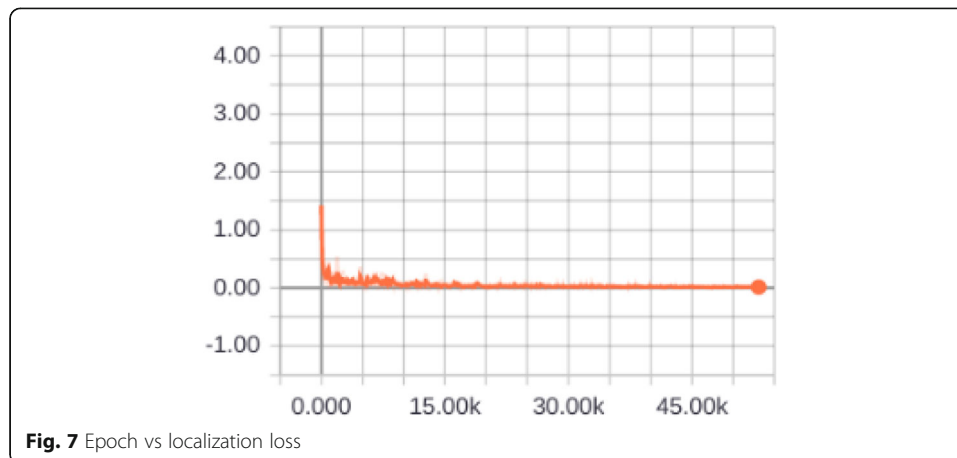
YOLOv3 architecture is suitable for small object detection because it uses a hybrid network.

4.1 Loss analysis

This section shows the evaluation of various losses such as total loss, classification loss, localization loss, clone loss, objectness, and localization loss in the region proposal network (RPN) for object detection. Figure 5 shows the total loss up to 45,000 epochs. It reached 0 from the 200 epoch that ensures very good accuracy. Figure 6 shows the classification loss that achieved 0 from the beginning of the epoch that ensures the perfect classification of the drones based on conditional probabilities. It is the final layer that reveals the object detection.

Figure 7 shows the localization loss during the various epochs. It returns the region proposals from the feature map based on the offset of the bounding box. The proposed





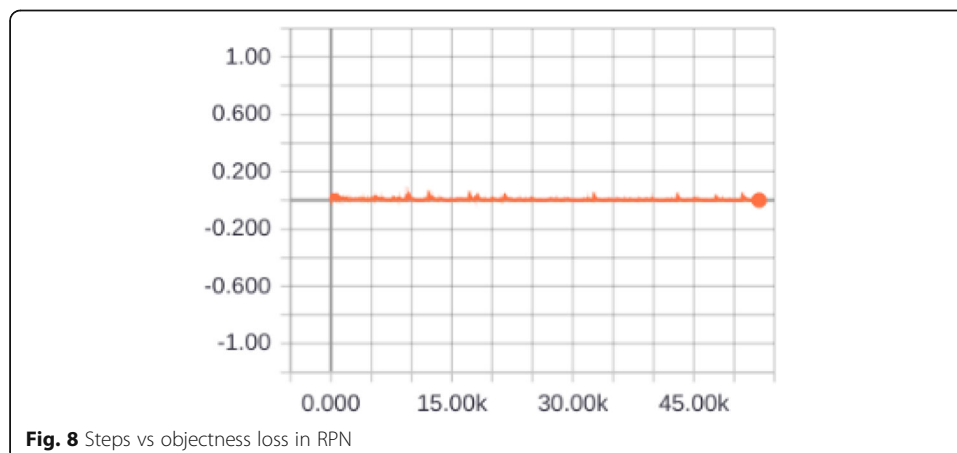
YOLOv3 scheme uses a sum-squared error (SSE) as loss function for optimization. It penalizes the classification error for an object in each grid cell. This scheme chose the box with the highest IOU.

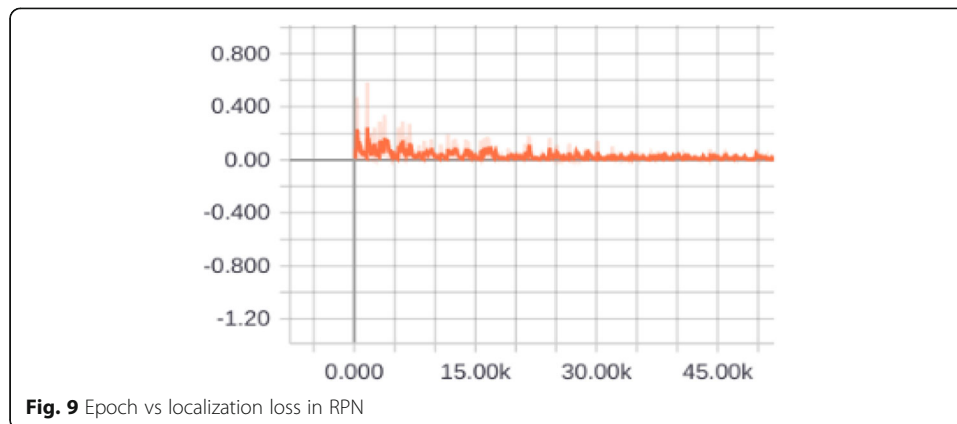
Figure 8 ensures that the proposed approach is extracting the region proposals for a targeted object. Region proposal network (RPN) is integrated with the convolution neural network CNN network for classification. The accuracy of the proposed scheme based on performance of the RPN module.

Figure 9 shows the localization loss in the RPN in which the bounding box regression loss is predicted to make good accuracy because it applied the regressor strategy for classification. Figure 10 shows the clone loss in various epochs. IT ensures the training and validation losses for accurate prediction. Also, it was recalculated by the neurons in the CNN layer weights for good predictions.

4.2 Accuracy

The proposed deep YOLOV3 model provides 99.99% accuracy while training and testing stage because the model is designed with 106 convolution layers and different size feature maps. The YOLOv2 model is also used for training and testing in which 98.27%

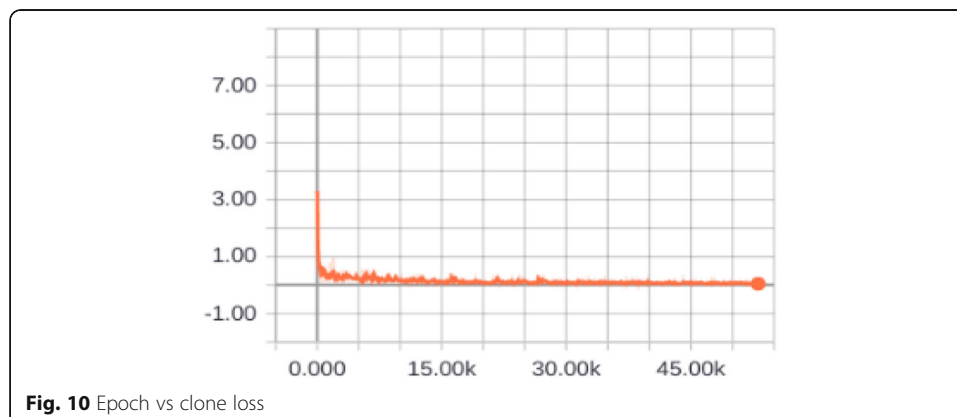


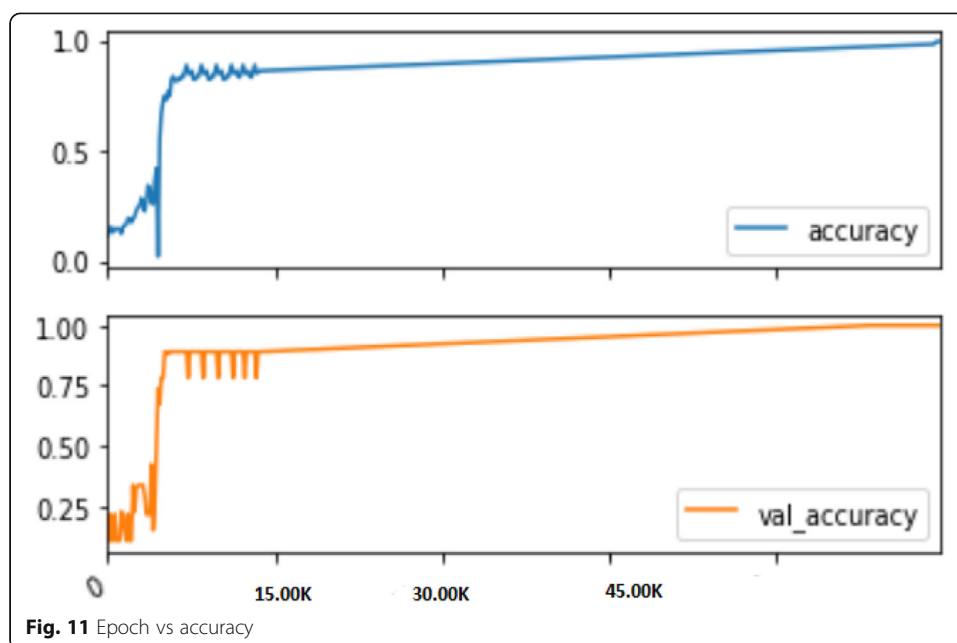


is achieved because it uses only residual network to detect an object. This model also applied a confidence score based on conditional probability to predict a target object effectively. Figure 11 shows that the proposed model achieved very good accuracy at the end of the epoch. The proposed approach also used backbone classifier to classify the objects accurately.

5 Conclusion

In this work, a novel deep YOLO V3 model is proposed to detect the small objects. The project involves training the model using pre-trained YOLOV3 with drone images. The simulation result is shown that the proposed deep YOLO V3 model is suitable for the computer vision process. In this, 106 convolution layers were designed with various feature maps to learn the small drone objects. YOLOv3 extracts better features using both Darknet and residual networks. The training stage performs 45,000 epochs to provide high accuracy. The proposed scheme uses IOU to predict a confidence score for a bounding box and grid cells simultaneously. The proposed model used logistic classifiers and binary cross-entropy loss for optimization to detect a small object. The proposed deep YOLO V3 revealed the 99.99% of accuracy because it used multi-scale predictions and backbone classifiers to better classify them. The different kinds of losses are analyzed that ensure the very good prediction of drone images because this model achieved a good





confidence score based on conditional probability. It is used to predict the accurate bounding boxes of an object. The proposed YOLOv3 model is not suitable to detect the larger size objects compared to previous versions such as YOLO and YOLOv2. In the future direction, the algorithm can be extended to train a huge volume of a small drone with the complex visible condition and far-flung remote areas.

Abbreviations

OSDDY: Object surveillance detection using deep YOLO; YOLO: You only look once; R-CNN: Region convolutional neural network; CBIR: Content-based image retrieval; VOC: Visual object classes; GPU: Graphics processing unit; VGGNet: Visual Geometry Group; UAV: Unmanned aerial vehicle; DCNN: Deep convolution neural network; RPN: Risk priority number

Acknowledgements

The authors would like to thank the anonymous reviewers for their helpful comments. The authors like to express special gratitude to Artificial Intelligence Lab, Ramco Institute of Technology, Tamilnadu, India, National Engineering College, Tamilnadu, India, and Sejong University, Seoul, South Korea, for providing and support for all the facilities to do this proposed experimentation.

Authors' contributions

M. Kaliappan: writing—original draft, writing—review and editing, S. Vimal: writing—original draft, conceptualization, data curation, validation. K. Vijayalakshmi: conceptualization, data curation, validation. Mi Young Lee: conceptualization, formal analysis, writing—review and editing, supervision, and funding. S. Manikandan: formal analysis, supervision, writing—review and editing. All authors read and approved the final manuscript.

Funding

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2018R1D1A1B07043302).

Availability of data and materials

Not applicable.

Declarations

Ethics approval and consent to participate

This research does not involve any human or animal participation.

Competing interests

The authors declare that they do not have any conflict of interests. All authors have checked and agreed the submission.

Author details¹Department of Computer Science and Engineering, Ramco Institute of Technology, Rajapalayam, Tamilnadu, India.²Department of Software, Sejong University, Seoul, South Korea.

Received: 10 May 2020 Accepted: 28 April 2021

Published online: 17 May 2021

References

1. S. Agarwal, J.O.D. Terrail, F. Jurie, *Recent Advances in Object Detection in the Age of Deep Convolutional Neural Networks*. *arXiv:1807.04606* (2018)
2. K. Muhammad, J. Ahmad, I. Mehmood, S. Rho, S.W. Baik, Convolutional neural networks based fire detection in surveillance videos. *IEEE Access* **6**, 18174–18183 (2018)
3. P.F. Felzenszwalb, R.B. Girshick, D. McAllester, D. Ramanan, Object detection with discriminatively trained part-based models. *IEEE Trans. Pattern Anal. Mach. Intell.* **32**(9), 1627–1645 (2010)
4. J. Zhang, K. Huang, Y. Yu, T. Tan, in *2011 IEEE Computer Vision and Pattern Recognition (CVPR), CO, USA, Colorado Springs*. Boosted local structured HOG-LBP for object localization (2011), pp. 1393–1400
5. Y. Zhang, S. Rho, S. Liu, D. Zhao, R. Ji, F. Jiang, 3D object retrieval with multi-feature collaboration and bipartite graph matching. *Neurocomputing* **195**, 40–49 (2016)
6. J. Zhang, Y. Huang, K. Huang, Z. Wu, T. Tan, in *Asian Conference on Computer Vision*. Data decomposition and spatial mixture modeling for part based model (2012), pp. 123–137
7. J. Park, S. Rho, C.S. Jeong, Real-time robust 3D object tracking and estimation for surveillance system. *Secur. Commun. Netw.* **7**(10), 1599–1611 (2014)
8. I. Krizhevsky, Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **25**, 1097–1105 (2012)
9. Y. LeCun, Y. Bengio, G. Hinton, Deep learning. *Nature* **521**(7553), 436–444 (2015)
10. E. Unlu, E. Zenou, N. Riviere, et al., Deep learning-based strategies for the detection and tracking of drones using several cameras. *IPSN Trans. Comput. Vis. Appl.* **11**, 7 (2019). <https://doi.org/10.1186/s41074-019-0059-x>
11. E. Mariappan, M. Kaliappan, S. Vimal, Energy efficient routing protocol using grover's searching algorithm for MANET. *Asian J. Inf. Technol.* **14**(24), 4986–4994 (2016). <https://doi.org/10.3923/ajit.2016.4986.4994>
12. G. Cao, X. Xie, W. Yang, Q. Liao, G. Shi, J. Wu, Feature-fused SSD: fast detection for small objects. *Comput. Vis. Pattern Recognit.* **10615**, 106151E (2018)
13. J. Ahmad, I. Mehmood, S. Rho, N. Chilamkurti, S.W. Baik, et al., *Comput. Electr. Eng.* **61**, 297–311 (2017). <https://doi.org/10.1016/j.compeleceng.2017.05.033>
14. J.H. Park, S. Rho, C.S. Jeong, J. Kim, Multiple 3D object position estimation and tracking using double filtering on multi-core processor. *Multimed. Tools Appl.* **63**(1), 161–180 (2013). <https://doi.org/10.1007/s11042-012-1029-9>
15. D. Kim, S. Rho, E. Hwang, Local feature-based multi-object recognition scheme for surveillance. *Eng. Appl. Artif. Intell.* **25**(7), 1373–1380 (2012)
16. S. Han, W. Shen, Z. Liu, *Deep Drone: Object Detection and Tracking for Smart Drones on Embedded System* (2016)
17. Y.C. Lee, J. Chen, C.W. Tseng, S.H. Lai, in *British Machine Vision Conference (BMVC)*. Accurate and robust face recognition from RGB-D images with a deep learning approach (2016)
18. J. Redmon, S. Divvala, R. Girshick, A. Farhadi, *You Only Look Once: Unified, Real-Time Object Detection*. *arXiv preprint arXiv:1506.02640* (2015)
19. W. Szegedy, Y. Liu, P. Jia, S. Sermanet, D. Reed, D. Anguelov, V. Erhan, Vanhoucke, A. Rabinovich, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Going deeper with convolutions (2015), pp. 1–9
20. K. He, X. Zhang, S. Ren, J. Sun, *Deep Residual Learning for Image Recognition*. *arXiv preprint arXiv:1512.03385* (2015)
21. J.F. Henriques, R. Caseiro, P. Martins, J. Batista, Highspeed tracking with kernelized correlation filters. *IEEE Trans. Pattern Anal. Mach. Intell.* **37**(3), 583–596 (2015)
22. F. Schroff, D. Kalenichenko, J. Philbin, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Facenet: a unified embedding for face recognition and clustering (2015), p. 815
23. D. Kim, S. Rho, E. Hwang, Real-time multi-objects recognition and tracking scheme. *Korean Navig. J. (KONI)* **16**(2), 386–393 (2012)
24. S. Hossain, D.-j. Lee, Deep learning-based real-time multiple-object detection and tracking from aerial imagery via a flying robot with GPU-based embedded devices. *Sensors (Basel)* **19**(15), 3371–3385 (2019). <https://doi.org/10.3390/s19153371>
25. T. Ojala, M. Pietikäinen, T. Mäenpää, Multiresolution gray scale and rotation invariant texture classification with local binary patterns. *IEEE Trans. Pattern Anal. Mach. Intell.* **24**(7) in press
26. R. Opromolla, G. Inchingolo, G. Fasano, Airborne visual detection and tracking of cooperative UAVs exploiting deep learning. *Sensors* **19**, 4332 (2019)
27. C. Kyrkou, G. Plastiras, T. Theocharides, *DroNet: Efficient Convolutional Neural Network Detector for Real-time UAV Applications*. *arXiv:1807.06789v1 [cs.CV]* (2018)
28. S. Vimal et al., Secure data packet transmission in MANET using enhanced identity-based cryptography. *Int. J. New Technol. Sci. Eng.* **3**(12), 35–42 (2016)
29. T. Lin, P. Goyal, R. Girshick, K. He and P. Dollár, "Focal Loss for Dense Object Detection," 2017 IEEE International Conference on Computer Vision (ICCV), 2999–3007 (2017). <https://doi.org/10.1109/ICCV.2017.324>
30. S. Annamalai, R. Udendhran, S. Vimal, in *Novel Practices and Trends in Grid and Cloud Computing*. An intelligent grid network based on cloud computing infrastructures (2019), pp. 59–73. <https://doi.org/10.4018/978-1-5225-9023-1.ch005>
31. S. Annamalai, R. Udendhran, S. Vimal, in *Novel Practices and Trends in Grid and Cloud Computing*. Cloud-based predictive maintenance and machine monitoring for intelligent manufacturing for automobile industry (2019), pp. 74–81. <https://doi.org/10.4018/978-1-5225-9023-1.ch006>
32. Y. Liu et al., *J. Phys. Conf. Ser.* **1345**, 062043 (2019)

33. L.I. Yundong et al., Multi-block SSD based on small object detection for UAV railway scene surveillance. *Chin. J. Aeronaut.* (2020). <https://doi.org/10.1016/j.cja.2020.02.024>
34. S. Vimal, M. Kaliappan, A. Suresh, P. Subbulakshmi, S. Kumar, D. Kumar, Development of cloud integrated internet of things based intruder detection system. *J. Comput. Theor. Nanosci.* **15**(11-12), 3565–3570 (2018)
35. J.-I. Watanabe, Y. Shao, N. Miura, Underwater and airborne monitoring of marine ecosystems and debris. *J. Appl. Remote Sens.* **13**(4), 044509 (2019). <https://doi.org/10.1117/1.JRS.13.044509>
36. M. Kaliappan, E. Mariappan, M. Viju Prakash, B. Paramasivan, Load balanced clustering technique in MANET using genetic algorithms defence science. *Def. Sci. J.* **66**(3), 251–258 (2016). <https://doi.org/10.14429/dsj.66.9205>
37. M. Kaliappan, B. Paramasivan, Secure and fair cluster head selection protocol for enhancing security in mobile ad hoc networks. *Sci. World J.* **2014**, 608984 (2014)
38. G.S. Kumar, M. Kaliappan, L.J. Julus, in *International Conference on Pattern Recognition, Periyar University*. Enhancing the performance of MANET using EESCP (2012)
39. S. Vimal, L. Kalaivani, M. Kaliappan, A. Suresh, X.-Z. Gao, R. Varatharajan, Development of secured data transmission using machine learning-based discrete-time partially observed Markov model and energy optimization in cognitive radio networks. *J. Neural Comput. Appl.*, 1–11 (2018). <https://doi.org/10.1007/s00521-018-3788-3>
40. A. Koksai, K.G. Ince, A. Alatan, in *Computer Vision and Pattern Recognition*. *arXiv preprint arXiv:2004.01059*. Effect of annotation errors on drone detection with YOLOv3 (2020) arxiv.org
41. S. Vimal, M. Khari, R.G. Crespo, L. Kalaivani, N. Dey, M. Kaliappan, Energy enhancement using Multiobjective Ant colony optimization with Double Q learning algorithm for IoT based cognitive radio networks. *Comput. Commun.* **154**, 481–490 (2020)
42. J.-H. Kim, Distortion invariant vehicle license plate extraction and recognition algorithm. *J. Korea Contents Assoc.* **11**(3), 1–8 (2011)
43. L. Zheng, C. Fu, Y. Zhao, in *International Conference on Digital Image Processing*. Extend the shallow part of single shot multibox detector via convolutional neural network (2018)
44. K. Simonyan, A. Zisserman, in *Int. Conf. Learn. Represent.* Very deep convolutional networks for large-scale image recognition (2015)
45. C. Szegedy et al., in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV*. Going deeper with convolutions (2015), pp. 1–9
46. S. Ioffe, C. Szegedy, in *International Conference on Machine Learning*. Batch normalization: accelerating deep network training by reducing internal covariate shift (2015), pp. 448–456
47. C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA*. Rethinking the inception architecture for computer vision (2016), pp. 2818–2826

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)
