e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:01/January-2022 Impact Factor- 6.752 www.irjmets.com

DEEP CONVONLUTIONAL NEURAL NETWORK FOR ABANDONED OBJECT DETECTION

Su Pyae Lwin*1, Myat Thida Tun*2

*1Department Of Computer Engineering And Information Technology, Yangon Technological University, Yangon, Myanmar.

*2Department Of Computer Engineering And Information Technology, Yangon Technological University, Yangon, Myanmar.

ABSTRACT

During the last few years, abandoned object detection in a video surveillance system has become essential for public safety. As a consequence, there has been a significant increase in researching the area of abandoned object detection. Some systems are not able to give an accurate output, some are not easy to implement and some are costly. Thus, the You Only Look Once (YOLO) approach was employed for object detection and Kalman Filter (KF) for object tracking. In this research, a review of YOLOv4 was extended by formalizing the framework for abandoned object detection. The network made predictions into six types of objects, person, backpack, handbag, book, umbrella, suitcase, by training a self-collected dataset. The experimental results gave a better performance for abandoned object detection by favoring robustness to errors such as illumination changes, a color match of foreground and background, and a high density of moving objects.

Keywords: Abandoned Object Detection, Video Surveillance, You Only Look Once, Yolov4, Kalman Filter.

I. INTRODUCTION

The government and big companies mostly used video surveillance a few years ago. But the use of surveillance cameras has gradually increased almost everywhere such as restaurants, offices, educational institutions and stations. Because the cost of technology is more reasonable and people take aware of a protection than a cure. But the criminals are also vigilant and observant. So, a key requirement to prevent unpredicted situations is the more advanced and accurate surveillance system. The requirement led to come up the idea of abandoned object detection through video surveillance [2]. Object detection has succeeded substantially with its high accuracy and available computing power by using deep learning. There are two types of deep learning-based object detection techniques such as two-stage and one-stage. In one of the basic two-stage object detectors, R-CNN, a selective search algorithm is firstly used to generate a large number of region proposals. In one-stage techniques, an image is looked at only once and it is not alike in two-stage techniques. YOLO is very popular in one-stage technique. YOLO means that only a single neural network evaluation is required to predict both class probabilities and bounding box coordinates for multiple objects in one image. YOLO makes more localization errors but less background errors than other state-of-the-art detection systems. The system is not able to stop working when unknown sources or unintended inputs are applied because objects are learned generally in YOLO [7]. On the Microsoft Common Objects in Context (MS COCO) dataset, 10% more mean Average Precision (mAP) has been achieved by YOLOv4 than the previous version, YOLOv3. Hence, in this research, objects are detected and classified by the YOLOv4 algorithm. Kalman Filter was employed to track moving or stationary objects because it can predict their future locations based on previous estimations even for the changing environment.

II. METHODOLOGY

There are various stages such as data pre-processing, training the model and post-processing to build an object detection model. But the new universal features in YOLOv4 are Weighted Residual Connections (WRC), Cross Stage Partial connections (CSP), Cross mini-Batch Normalization (CmBN), Self Adversarial Training (SAT), Mish activation, Mosaic data augmentation, DropBlock regularization, and CIoU loss in combination to achieve high average precision (AP) and frames per second (FPS). A one-stage detector architecture is followed by YOLOv4 composing of four parts such as input, backbone, neck, and dense prediction or head. In the input, the dataset is required to be feed for detection. For the extensible and strong object detector, the backbone enables to extract features and use the image dataset. There are three parts in the backbone such as bag of freebies (BoF), bag of

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:01/January-2022 Impact Factor- 6.752 www.irjmets.com

specials (BoS), and CSPDarknet53. Deep high level features are extracted and class probability scores are predicted in the backbone. The head works in the same way as YOLOv3. In traditional object detection algorithms, the bounding boxes are refined, similar detections are removed, and the box scores are redefined depending on other objects in the view as post-processing. But YOLOv4 uses non-maximum suppression. Figure 1 showed that the high probability boxes are taken and the boxes with high Intersection over Union (IoU) are suppressed. It repeats these two functions until a box is selected and considers that as the bounding box for that object [1][3].

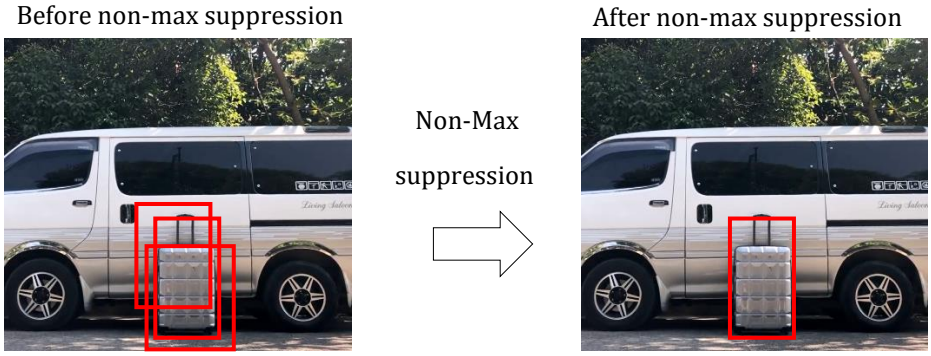

Figure 1: Non Max Suppression

Kalman Filter

A Kalman filter mainly executes to estimate the state which is distributed by a Gaussian of a linear system. The filter is recursive and predictive because it depends on the application of recursive algorithms and state space techniques. Hence, the state of the system is dynamically estimated. White noise or some noise can mostly disturb this dynamic system. Measurements are used by the filter to enhance the estimated state but the state is able to be disturbed a little. Thus, Kalman filtering comprises two stages which are prediction and correction [4]. When the system has a person detection, person tracking is started and a new object is waited to appear right after the existence of the person. Figure 2 showed that object tracking is continued and unalarmed waiting for the non- existence of the person. When the person is left, the person is waited to return for more than 20 seconds and checked every frame. Figure 3 showed that just after 20 seconds of the non-existence of the person, the abandoned object is alert by drawing a bounding box even if any motion exists in the surveillance camera.[6].

Figure 2: Attended Backpack

Figure 3: Unattended Backpack

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:01/January-2022 Impact Factor- 6.752 www.irjmets.com

III. MODELING AND ANALYSIS

In this study, one of the state-of-the-art approaches, YOLOv4, was chosen for object detection because it is able to detect in real-time with high speed, and accuracy. To continue the network training, 3329 images consisting of images of 562 people, 559 backpacks, 552 handbags, 564 books, 544 umbrellas, 548 suitcases, were collected from different kinds of on ground captures and Google. The dataset is the ratio of 90 to 10 for train_test_split. For image annotation, LabelImg tool was used and the images were manually annotated using classes; person was denoted as the "zero class", backpack the "first class", handbag the "second class", book the "third class", umbrella the "fourth class" and luggage the "fifth class". For the YOLO implementation, the annotated images were saved in the format of .txt. Then, the YOLOv4 architecture was configured and fine-tuned for the custom dataset. There were fine-tuning of the final three YOLO and convolutional layers for a specific number of classes. Thus, the total classes were specified as six, namely "person", "backpack", "handbag", "book", "umbrella" and "suitcase". Before each of the three layers of YOLO, the number of convolutional layers were three in order to build a high-level feature map which is robust to illumination changes, a color match of foreground and background and partial occlusion as shown in Figures 4,5 and 6.

Figure 4: Unattended Illumination Changed Handbag

Figure 5: Unattended Color-matched Handbag

Figure 6: Unattended Occluded Handbag

Features are extracted by using filters in convolutional layers. The total number of filters can be achieved by the calculation of (number of classes + 5) × 3. Hence, 33 was obtained for the filter number of the three convolutional layers ahead of the YOLO layers. The remaining layers were kept implementing through the similar 162 layers. The several approaches for data augmentation highlighted the issues of data scarcity in YOLOv4. The MOSAIC flag was switched on to execute the automated data augmentation [8]. The batch number was defined to 64. The numbers that are tried for subdivisions are various and went from 8 to a factor of 8 based on the GPU. In the research, subdivision = 32 was appropriate. The image was set to the size of 608 × 608

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:01/January-2022 Impact Factor- 6.752 www.irjmets.com

for width and height. The default values were 0.001 for learning rate, 0.949 for momentum, 0.0005 for decay, 0.1 for hue, 1 for normalization and mish for activation in other hyperparameters. Further, the max batch number was fine-tuned and set to 12000, that was achieved by the calculation of (number of classes × 2000). The percentages of eighty and ninety of the max batches defined the steps. Thus, the step was set in the range of 3200 and 3600. Finally, the YOLOv4 was trained on Google's DL virtual machine and the continuous testing of images and videos were then performed. YOLOv4 was trained for 12000 iterations and the trained weights were accumulated at the iterations of 1000, 2000, 3000, and 12000 by the open source neural network framework, Darknet. After the completion of object tracking by Kalman Filter, the system assumed that the object is abandoned if the owner is found to be missing from the camera scene for longer than t seconds [5]. The final output was alert by bounding boxes as shown in Figure 7.

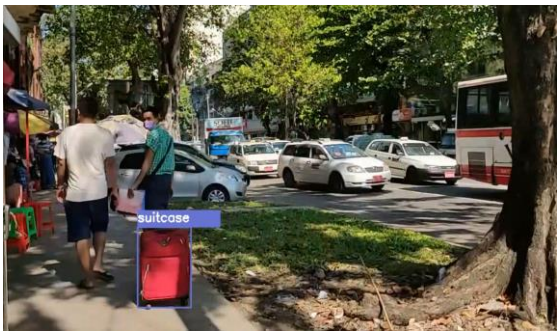

Figure 7: Abandoned Suitcase Detection

IV. RESULTS AND DISCUSSION

The experiment was implemented employing the Darknet framework and Google's DL virtual machine. A Tesla with K80 was applied as Graphic Processing Unit (GPU) for training the Darknet. A cuDNN with the version of 7.6.5 was set up for running programs on the NVIDIA GPU. The total amount of time required to train the network with the above configurations was approximately 2 days. The output model of YOLOv4 performed very well according to the key indicators such as precision, recall, F1-scores, and mAP. The values of the indicators can be seen in Table 1. During 12000 iterations, 84.55 % was achieved as the highest mAP. In addition, the twenty captured videos were collected in four different scenarios such as indoor, outdoor, night and crowded scene. Table 2 showed that almost all objects were truly detected for video detection in a complex background and various types of weather such as sunny, windy and cloudy except that fully occluded objects in crowded scene were not able to be detected.

Table 1. Performance evaluation of the experiment

Precision	Recall	F1-Score	mAP
0.79	0.86	0.83	84.55%

Table 2. Abandoned object detection for four different scenarios

Method	Video Clips	Ground Truth	True Detections	False Detections	Scenario
YOLOv4	Video_1.mp4	1	1	0	Outdoor
	Video_2.mp4	1	1	0	Outdoor
	Video_3.mp4	1	1	0	Outdoor
	Video_4.mp4	1	1	0	Outdoor
	Video_5.mp4	1	1	0	Outdoor
	Video_6.mp4	1	1	0	Outdoor
	Video_7.mp4	2	2	0	Indoor
	Video_8.mp4	1	1	0	Indoor

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:01/January-2022 Impact Factor- 6.752 www.irjmets.com

	Video_9.mp4	1	1	0	Indoor
	Video_10.mp4	1	1	0	Indoor
	Video_11.mp4	1	1	0	Indoor
	Video_12.mp4	1	1	0	Night
	Video_13.mp4	1	1	0	Crowded Scene
	Video_14.mp4	2	1	1	Crowded Scene
	Video_15.mp4	1	1	0	Crowded Scene
	Video_16.mp4	1	1	0	Crowded Scene
	Video_17.mp4	1	1	0	Crowded Scene
	Video_18.mp4	1	1	0	Crowded Scene
	Video_19.mp4	1	1	0	Crowded Scene
	Video_20.mp4	1	1	0	Crowded Scene

V. CONCLUSION

This research presented an abandoned object detection system for video surveillance which is reliable, interactive and effective in terms of processing speed and accuracy by using YOLOv4. For further extension, a larger scaled image dataset with multiple different classes will be created for high-level semantic descriptions of abandoned objects. In addition, the other trainable and comparable object detection algorithms are R-CNN, mobilenet, retinanet, etc. YOLOv5 has also been released. Thus, this version will further be used to increase accuracy and decrease the execution time.

ACKNOWLEDGEMENTS

I would like to thank Associate Professor Myat Thida Tun for her guidance and support and I also would like to thank Head of Department Dr. Khine Thinzar for letting me write this journal.

VI. REFERENCES

[1] Joseph Redmon, Santosh Divvala, Ross Girshick and Ali Farhadi , “You Only Look Once: Unified, Real-Time Object Detection,” submitted for publication, 2016.

[2] YingLi Tian, Rogerio Feris, Haowei Liu, Arun Humpapur, and Ming-Ting Sun, “Robust Detection of Abandoned and Removed Objects in Complex Surveillance Videos”, submitted for publication, 2017.

[3] Meenal Bargath, Monika Verma, Ritika, Shalini Vishwakarma, Ruchi Biswas, Shailendra Singh,“Abandoned Object Detection”, International Journal of Advanced Research in Computer Science and Software Engineering, ISSN: 2277 128X, Volume 7, Issue 4, April 2017.

[4] Mandakini A. Mahale, H. H. Kulkarni, “Abandoned Object Detection with Video Surveillance using Gaussian Mixture Model and Kalman Filter”, ISBN: 978-93-86171-81-8 submitted for publication, on 26 Nov 2017.

[5] Jovin Angelico and Ken Ratri Retno Wardani “Convolutional Neural Network Using Kalman Filter for Human Detection and Tracking on RGB-D Video”, CommIT (Communication & Information Technology) Journal 12(2), 105–110, 2018.

[6] Su Su Aung, Nay Chi Lynn, “A Study on Abandoned Object Detection Methods in Video Surveillance System,” University of Computer Studies, Mandalay, Myanmar, submitted for publication, 2020.

[7] Omkar Masurekar, Omkar Jadhav, Prateek Kulkarni, Shubham Patil, “Real Time Object Detection Using YOLOv3,” International Research Journal of Engineering and Technology (IRJET), Volume: 07, Issue: 3, Mar 2020.

[8] Alexey Bochkovskiy, Chien-Yao Wang and Hong-Yuan Mark Liao, “YOLOv4: Optimal Speed and Accuracy of Object Detection,” arXiv:2004.10934v1, submitted on 23 Apr 2020.