ISSN 2320 - 2602
Volume 10 No.3, March 2021
International Journal of Advances in Computer Science and Technology
Available Online at http://www.warse.org/IJACST/static/pdf/file/ijacst011032021.pdf
https://doi.org/10.30534/ijacst/2021/011032021

Video Object Detection for Police Surveillance using Deep Learning

Marco Marvin L. Rado[1], Maria Visitacion N. Gumabay[2]
1 St.Paul University Philippines, Philippines, mlrado@liceo.edu.ph
2St.Paul University Philippines, Philippines, mvgumabay@spup.edu.ph

ABSTRACT

Human vision is incredibly excellent and complex. In the previous years, people made significantly more leaps to expanding this visual capacity to machines. Cameras have been used as the eyes of computers.

In response to increasing anxieties about crime and its threat to security and safety, the utilization of substantial numbers of closed-circuit television system (CCTV) in both public and private spaces have been considered a necessity. The use of these significant video footages is essential to incident investigations. But as the number of these systems rises, so as the need for human operator monitoring tasks. Unfortunately, many actionable incidents are utterly undetected in this manual system due to inherent limitations from deploying solely human operators eye-balling CCTV screens. As a result, surveillance footages are often used merely as passive records or as evidence for post-event investigations.

This study aimed to develop a real-time firearm detection using deep learning embedded in CCTV cameras that pushes alert notifications to both iOS and Android mobile devices. This research used a descriptive design and asked IT experts to evaluate the develop system based on its compliance to ISO 25010 standard. Moreover, confusion matrix and intersection over union (IoU) were used to evaluate the performance of the system. The detection system was found to be highly recommended in urban areas particularly for CCTVs found in barangay streets and establishments.

Key words: Android, CCTV, Computer Vision, Faster RCNN, Firebase, iOS, Python, SSDLite

1. INTRODUCTION

The event of robbery, harassment, vandalism and other wrongdoings has increased generally on numerous establishments in the Philippines. It is viewed as disturbing since such occurrences have been executed in broad daylight, and majority of them are happening late at night up to the wee hours in the morning. Subsequently, the Philippine National Police (PNP) considered the installation of closed-circuit television system (CCTV)s or similar surveillance digital services in addressing the rising crime incidence in the society particularly in the urban side, where in the use of any of these significant video footages that has been deemed essential or is subject to a request by any entity or is subject to investigation is institutionalized. Moreover, these are installed inside business establishments, entrances, and exits that include pathways and streets for passerby and parking lots. These cameras record videos and retain continuous digital images for review and reference purposes. At the same time, authorities are actively looking forward to creating safety mechanisms that assist in protecting people and establishments by making criminal offenders feel: (a) They may be unable to commit crimes as a result of an increased punishment certainty generated from increased CCTV detections and subsequent enforcement actions carried by responding patrol officers [6] and (b) They might find themselves in a situation where they may be apprehended, which will cause in them being reluctant to do crimes where such systems are in place. As stated in the study of Piza *et al*. [6], of the 74 reported total incidents to patrol officers, sixty-four (56.5%) of the detections resulted in an enforcement action. Thirty-nine of the 64 enforcement actions (60.9%) were arrests. The remaining 25 (39.0%) were record checks or field interrogations.

Computer vision is making sense of what a machine sees. Applying deep learning in object detection, it teaches machines how to perform a specific task given a set of examples. Over the past years, deep learning methods have been shown to outperform previous state-of-the-art machine learning techniques in several fields[7,8].

In this study, the proponent incorporated mobile devices both in Android and iOS in CCTV systems with detection mechanism to alert police officers through notifications of a detected firearm in a specific location of the CCTV camera feed. There are similar systems in the market but there are no products that are very customizable and trainable like the deep learning framework's capability to be scaled to future

needs. This research gathered images of pistols, rifles, knifes, police and security personnel that are commonly present in crime incidents.

2. LITERATURE REVIEW

Detection is the process of observing or finding something. At the terminal, there are dogs trained in the detection of drug trafficking or explosive materials in baggage. In the military, detection refers to the discipline of observation of a region to locate an enemy or ascertain strategic features. In computer technology related to computer vision and image processing, object detection is the process of finding instances of real-world objects of a specific class such as humans, buildings, or cars in digital images and videos. Object detection is a fundamental visual recognition problem in computer vision, and it is widely studied in the past decades [8].

The proponent developed a CCTV deep learning and mobile based application to help police personnel deter crimes or apprehend offenders by detecting firearms in CCTV feeds and push notifications in mobile devices. It helps to identify the object of interest and the location of the detection. The system can also have a significant impact on the way information is collected, stored, and accessed for further data analytics of crime related information.

The breakthrough of deep learning was seen by Deng et al. [1] when a collection of a large-scale annotated image dataset ImageNet, which contained 1.2 million high resolution images was done, making this technology to train deep models with large scale training data. There is a proven advantage in training with large available data sets, as this will improve the performance of the system. With the development of computing resources on parallel computing systems such as GPU clusters, in 2012, a large deep convolution model was trained with the ImageNet dataset and showed significant improvement on Large Scale Visual Recognition Challenge (ILSVRC) compared to all other approaches [2]. After the success of applying Deep Convolutional Neural Network for classification, deep learning techniques were quickly adapted to other vision tasks and showed promising results compared to the traditional methods. Currently, deep learning-based object detection frameworks can be primarily divided into two families: (a) two-stage detectors, such as Region-based Convolutional Neural Network (R-CNN) which was employed in this study, specifically Faster R-CNN for the computer desktop deployment and training, and its variants [3]; and (b) one-stage detectors, such as You Only Look Once (YOLO) and its variants, Single Shot Multibox Detector (SSD) [4] which was employed for the Raspberry Pi deployment in this study.

3. METHODOLOGY

This paper used the descriptive method for the evaluation of the developed system based on ISO standards. The ISO 25010 are needed to evaluate software systems [5]. Rapid Application Development (RAD) was also used for the development of the detection system and mobile applications.

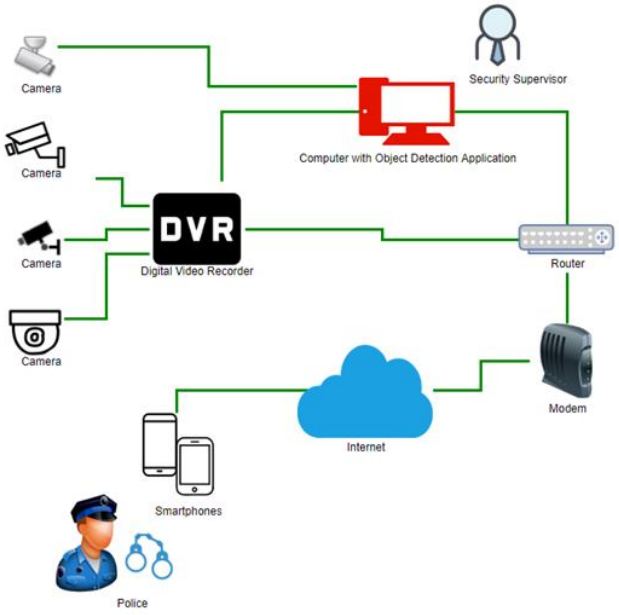

Figure 1: Detection System Architecture

Figure 1 illustrates the overall system architecture. The significant video footage from the police investigators was manually annotated and was fed to train the system's detector. To speed up the training process, a pre-trained model was employed. The knowledge gained from a pre-trained model enhanced the performance of the system's detector. The trained model was tested under real-time setting using selected CCTV with Digital Video Recorder and additional web cameras from the same location. Once there is a detection made, a developed Python function will push a notification to nearby police iOS and Android smartphones using Firebase Cloud Messaging (FCM) about the detection and its location using the ID of the CCTV camera.

4. RESULT AND DISCUSSION

The developed was tested using: (1) Confusion matrix for the video samples; and (2) Intersection over union (IoU) for still images.

n=38	Predicted No	Predicted Yes	
Actual No	TN 38	FP 0	
Actual Yes	FN 7	TP 31	Recall $\frac{TP}{TP+FN} = \frac{31}{31+7} = 0.82$
	Precision $\frac{TP}{TP+FP} = \frac{31}{31+0} = 1$	Accuracy $\frac{TP+TN}{TP+FP+TN+FN} = \frac{31+38}{31+0+38+7} = 0.91$	

Figure 2: Confusion Matrix Results of the Detection System

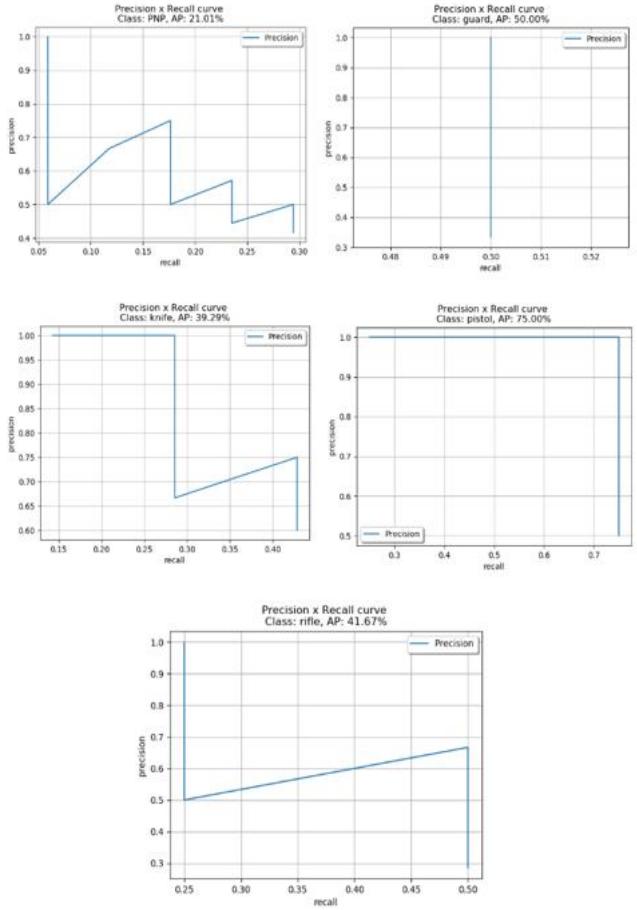

Figure 3: Precision Averages of the Objects using Precision x Recall Curve at 0.7 threshold

In this study, the detection system was successful with promising results. In figure 2, the detection system got 91% in accuracy, 100% in precision, and 82% in recall when tested with 38 video samples.

The average precision of the five objects of interest of the detection system is shown in figure 3. The police, security guard, knife, pistol, and rifle class got 21%, 50%, 39%, 75%, and 42% average precisions respectively using the collected images for training testing. Overall, the mean average precision (mAP) is 45.37% using the object detection metrics, see figure 4.

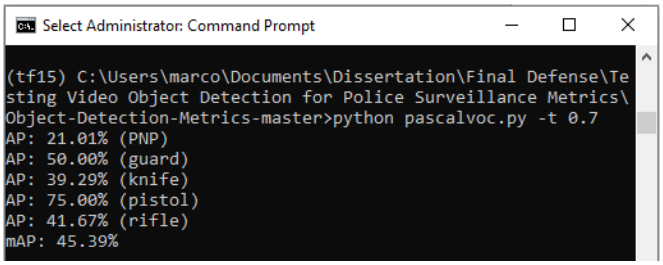

Figure 4: Mean Average Precision (mAP)of the Detection System using the Object Detection Metrics at threshold 0.7

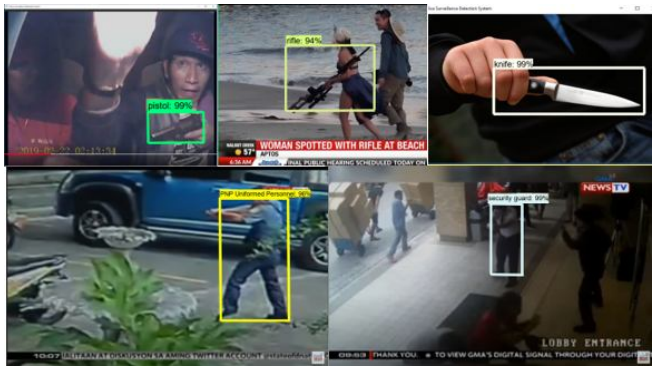

Figure 5: Detection System as viewed in Computer Monitors

The metrics were evaluated in comparison to the ground truth data. These ground truth data were from the training, validation, and test datasets. The ground truth included the image, the classes of the objects in it, and the true bounding boxes of each of the objects in the image.

During the evaluation process, it run the original image through the trained model and results were returned and calculated after confidence thresholding.

In calculating precision and recall, true positives, false positives, true negatives, and false negatives were identified. The COCO evaluation metric recommended measurement across various IoU thresholds at 0.5, which is also the PASCAL VOC metric. But for this study, the threshold was 0.7, true positive is considered only if IoU is greater than 0.7 else it is considered a false positive.For recall, negatives were counted. Every part of the image where it did not predict an object is considered negative. But only false negatives were measured, that is, the objects that the model has missed. True negatives were not measured as it is futile.

Figure 5 shows the detection of a pistol, rifle, knife, police officer, and security personnel with its corresponding percentages in the computer monitors.

Figure 6 shows the alert notifications as received in both iOS and Android mobile devices. The details include the

mobile app name, the time when it is received, the classification of the detection object and its corresponding percentages, and the specific location of the detection as determined by the location of the CCTV camera.

Figure 6: The Alert Notifications as displayed in both iOS and Android mobile devices

Table 1 presents the summary of the assessment of the IT experts on the detection system's compliance to ISO 25010 software quality standards. Overall, it can be observed in the data that the developed system complied to the ISO 25010 software quality standards to a "great extent" with an overall mean of 4.06.IT experts evaluated the detection system's functional suitability when the video feed transitioned frame by frame giving this a descriptive interpretation of moderate extent. Furthermore, the system's performance efficiency got a moderate extent as it requires higher computing power to deliver smooth video monitoring in computer monitors. Finally, the system got a moderate extent it terms of its usability as they experienced intricacies in its installation, configuration, training, and deployment.

Table 1: Summary of Assessment of the IT Experts of the Detection System to all 8 ISO 25010 Standards

Category	Mean	Descriptive Interpretation
Functional Suitability	3.32	Moderate Extent
Performance Efficiency	3.36	Moderate Extent
Compatibility	3.55	Great Extent
Usability	3.32	Moderate Extent
Reliability	3.91	Great Extent
Security	4.00	Great Extent
Maintainability	3.51	Great Extent
Portability	4.00	Great Extent
Overall Mean	**4.06**	**Great Extent**

5. CONCLUSION

Based on the results obtained from the study, the following conclusions were drawn:

Promising results were obtained when deep learning technology is integrated with CCTV systems, firebase cloud messaging and mobile applications. This creates a utility value that could benefit the community. The given assessment of the IT experts to the developed system proved that it could help in improving the objective of the police to somehow deter crime and apprehend offenders in real time.

REFERENCES

1. J. Deng, W. Dong, R. Socher, L.J. Li, K. Li, L. Fei-Fei. **ImageNet: A large-scale hierarchical image database**, *IEEE Conference on Computer Vision and Pattern Recognition*, Miami, FL, USA, 2009
2. A.Krizhevsky, I.Sutskever, G. Hinton. **Image classification with deep convolutional neural networks**, in: *NeurIPS*. 2012
3. T.Lin, P.Dollar, R.Girshick, K.He, B.Hariharan, S.Belongie.**Feature pyramid networks for object detection**, in: *CVPR*. 2017
4. W.Liu, D.Anguelov, D.Erhan, C.Szegedy, S.Reed, C.Fu, A.Berg.**SSD: Single Shot Multibox Detector**, in: *ECCV*. 2016
5. E.Peters, andG. K.Aggrey.**Evaluating the Effectiveness of ERP Systems in HEIs: A Proposed Analytic Framework,***International Conference on Computing, Computational Modelling and Applications (ICCMA)* (pp. 40-45). IEEE.2019
6. E.Piza, J.Caplan, L.Kennedy, and A.Gilchrist. **The Effects of Merging Proactive CCTV Monitoring with Directed Police Patrol: A Randomized Controlled Trial.***Journal of Experimental Criminology*, 11(1): 43-69.2015
7. S.Ren, K.He, R.Girshick, and J.Sun. **Faster R-CNN: Towards real-time object detection with region proposal networks**, *in: NeurIPS*. 2015
8. X. Wu, D. Sahoo, and S. Hoi. **Recent Advances in Deep Learning for Object Detection.**arXiv:1908.03673. Retrieved fromhttps://arxiv.org/abs/1908.036732019