Object Detection: Harmful Weapons Detection using YOLOv4

Wan Emilya Izzety Binti Wan Noor Afandi
School of Electrical Engineering
College of Engineering
Universiti Teknologi MARA
40450 Shah Alam, Malaysia
emilyaizzety@gmail.com

Naimah Mat Isa
School of Electrical Engineering
College of Engineering
Universiti Teknologi MARA
40450 Shah Alam, Malaysia
naimahmatisa@gmail.com

2021 IEEE Symposium on Wireless Technology & Applications (ISWTA) | 978-1-6654-4043-1/21/$31.00 ©2021 IEEE | DOI: 10.1109/ISWTA52208.2021.9587423

Abstract— **Closed-circuit television (CCTV) is essential in the security industry by providing surveillance, monitoring activities, recording incidents, and storing evidence. Research and developments have been performed to ameliorate its application to meet the ever-changing security landscape. This paper presents a revolutionary method to enhance the application of CCTVs in Malaysia. The purpose of this study is to develop an Artificial Intelligence (AI) based weapons detection that helps people in identifying violent crimes that are currently happening. This study focuses on detecting harmful weapons such as handguns and knives using the custom trained object detection model that has been trained using the YOLOv4 Darknet framework. Two sets of training have been done to test the effectiveness of this system. The first training was done on a single class custom object detection model while the second was done on a multiple class custom object detection model. Based on the results obtained, the single class object detection only managed to achieve 66.67% to 77.78% accuracy on average whilst the multiple class object detection managed to achieve up to 100% accuracy on most of its input images.**

Keywords— CCTV; Artificial Intelligence (AI); Object detection; Weapon; YOLOv4

INTRODUCTION

The most primitive documented usage of closed-circuit television (CCTV) technology was first used in Germany [1], the system was set up to monitor the V-2 rockets [2]. The world's first CCTV can only be used for live monitoring and not to record footage [2]. Not long after, the system was being promoted by a vendor called Vericon [3] and has been made available to the public commercially. This technology has significantly improved over the years and the recording systems even became more versatile and dependable as seen today.

As is known, conventional CCTV requires constant monitoring by security personnel. The drawbacks of conventional CCTV are that security personnel might miss incidents happening within the background as they are focusing on something more prominent. Apart from that, excessive usage of screen time can also lead to an eye problem called computer vision syndrome [4] showing symptoms such as redness, dryness, blurred vision, and double vision.

On top of that, CCTV that relies on motions only has a high possibility of giving out false alarms that will be a disadvantage to CCTV systems. A study has shown that spiders and their cobwebs cause some of the most popular false alarms [5] that an operator has to deal with apart from pets. A method to detect a suspicious person today is by detecting a person idling around at a certain location [6]. This method is however inaccurate as some might just be waiting for friends or family.

To curb this issue, the application of Artificial Intelligence (AI) can be implemented. This study is proposed to help reduce violent crimes from happening. This system provides a solution for regular CCTV by detecting two types of weapons i.e., Handguns and Knives. It will take the security industry to a new level and hopefully see a declining statistics in armed crimes. This system is meant to be implemented on CCTV or alarm security systems for security purposes but for this study, it is only limited to the detection of the object. Further steps has to be undertaken before the system can be completely applied in a CCTV security system.

Apart from that, a study on the mean average precision (mAP) of two custom trained object detection models has been made to compare the differences of both results obtained. Lastly, to test the effectiveness of this system, accuracy is calculated and compared for each of the input images.

LITERATURE REVIEW

Previous researchers have done studies on how to protect society from violent crimes. One of the most focused areas is the versatility of CCTV. A few studies have reported on the effectiveness of CCTV implementations. Some did not have any effect after the installation of CCTV [7] while others show a significant reduction in violent crimes [8]. The following research listed below will be the primary source of ideas and motivation in the establishment of this project.

AI-Based Automatic Robbery/Theft Detection using Smart Surveillance in Banks

Their focus is on implementing a Smart Cam that monitors the bank's activity. Their system can detect any kind of suspicious behaviour. This Smart Cam can also detect the types of weapons and count the number of weapons it has detected. Once a weapon or suspicious behaviour has been detected, the thieves would be tracked, and an alert notification will be sent automatically including the details as shown in Fig. 1 to the security department [9].

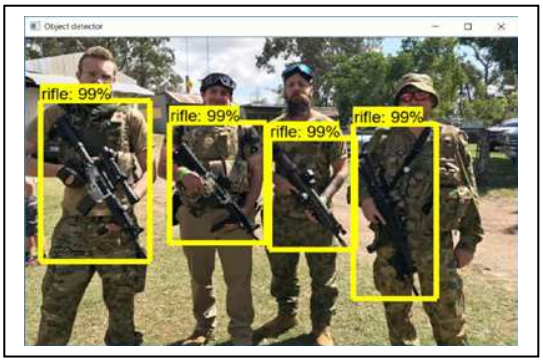

Fig. 1. Object detected after training. [9]

A Neural Network Based Intelligent Intruders and Tracking System using CCTV Images

Early development on intruder detection and tracking system based on the neural network approach has been made. According to this paper, the potential intruder is identified by examining the technique and algorithm in a neural network. When the system identifies the presence of an intruder, it will start monitoring and tracking its movement. This way, the information gathered can be used for further identification. Apart from that, they made a comparison between the traditional approach of Intelligent Scene Monitoring (ISM) and the artificial neural network (ANN). It is said that the ANN approach can differentiate between suspicious behaviour and non-suspicious behaviour [10].

YOLOv4: Optimal Speed and Accuracy of Object Detection

A study has been made to compare different types of object detectors. As for their graphics processing unit (GPU)s, only three types of GPU are used for this comparison which is Maxwell, Pascal and Volta. According to their results, the proposed YOLOv4 are located at where they called the 'Pareto optimality' curve. This means in terms of speed and accuracy, their object detector is the fastest and the most accurate as compared to other available detectors. Based on their findings, YOLOv4 was proven to produce improved results towards YOLOv3 by 10% while for FPS, it improves up to 12% [11].

METHODOLOGY

This study proposed a weapons detector framework to detect the presence of harmful weapons, by analyzing the image or video frame by frame. The purpose of detecting weapons is because incidents involving the use of firearms remain a major threat to national security. Although armed robbery in Malaysia has decreased significantly [12] but appropriate action should be taken to make sure armed robbery in Malaysia is under control. Hence, this will be a breakthrough in security management as potential crimes involving firearms can be effectively curtailed.

The proposed idea is by using Darknet YOLOv4 and TensorFlow platform. Darknet is an open-source library to build a neural network framework while TensorFlow is a platform to run the YOLOv4 detector. Further explanation will be explained in this section.

(1) Dataset Preparation

Instead of using a CCTV video, this research is done offline and by using an images dataset. The images came from the Open Images V6 dataset. Open Images V6 dataset consists of 9.6 million images with annotations for segmentation, object detection and classification process [13]. The images are downloaded with the label for training purposes. More than 3000 images are downloaded and used in this project. Then, the images are sorted and labelled according to the YOLOv4 format.

(2) Training YOLOv4

For this project, the training session was done by using Google Colab. The YOLOv4 is trained to detect harmful weapons. The training dataset is saved in google drive where the dataset can be retrieved by google Colab. The training session is divided into two sessions as the first session is to train single class object detection and the second session is multiple class object detection. Before the training session is started, some parameters need to be defined i.e. batch size, subdivisions, maximum batches, number of classes, width and height.

(3) Object Detection Model

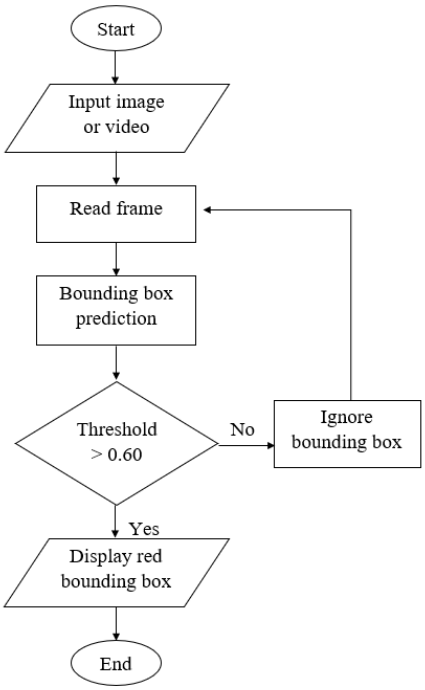

Fig. 2. Single Class Object Detection Model for AI Based Monitoring System flowchart.

The system flowchart of a single class object detection model is shown in Fig. 2. First, the process starts by reading the input image or video footage frame by frame but for this project an image is used. Then the model starts to detect an object on the input images. The detected object is bounded with the bounding box where the bounding box has a threshold value to be achieved. In this system, the threshold value is set to a minimum of 0.6 therefore when the threshold is above 0.6, only then it will display a red bounding box. Or

else, it will just ignore the predicted bounding box and continue to read frames.

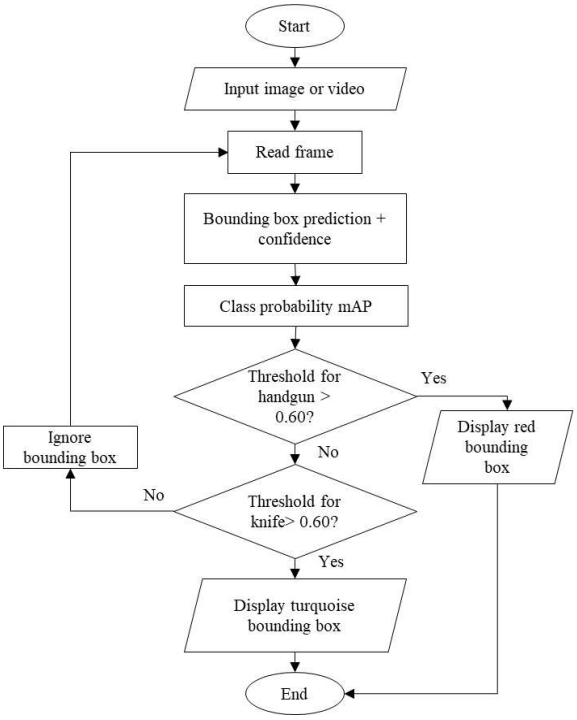

Fig. 3. Multiple Class Object Detection for AI Based Monitoring System flowchart.

The process flow in Fig. 3 is slightly different as compared to Fig. 2 because there are two types of objects that needs to be detected. It starts by reading the input image or video footage frame by frame. The input is splitted into grid cells to predict each bounding box simultaneously along with its confidence value. The confidence score indicates how confident the model is when predicting it. Each of the grid cells also predicts the class probability that it belongs. The predicted class will not be displayed if the confidence score is less than 0.6. This conditions also applies to the process as depicted in Fig. 2. Therefore, each predicted class will be evaluated. If the threshold for a class called handgun is above 0.6, a red bounding box will be displayed. Next, it will check if the threshold for the class called knife is above 0.6, if the result is true, a turquoise bounding box will be displayed. Or else, it will just ignore the predicted bounding box and continue to read the frame.

Open Images V6 Dataset

Open Images v6 is the dataset used to train the model. The images collected are more than 3000 images with the label. The image size varies and before training the default value for image size is 416 x 461 pixels.

Training YOLOv4 custom Object Detector

In this study, all training were done in the cloud using Google Colab. Colab's graphics processing unit (GPU) also includes Nvidia K80s, T4s, P4s and P100s which has better GPU performance compared to any laptop's GPU. The YOLOv4 is the new version of the state-of-the-art You-Only-Look-Once real-time object detection algorithm trained in the Darknet framework. It is chosen due to its speed and accuracy. This makes training so much faster even on a single GPU.

Object Detection Model

The YOLOv4 is implemented on TensorFlow. The TensorFlow framework is only used to run the object detection model. It is chosen due to its flexibility as it is a low-level library. It can be changed according to the requirements needed. Before the TensorFlow framework can be used, the YOLOv4 weights file has to be converted to a TensorFlow PB file.

RESULT AND DISCUSSION

This section compares the results of object detection when it was executed using a single class versus multiple classes. Object detection model was trained using the YOLOv4 algorithm in the Darknet framework for custom datasets of a single class and datasets of two classes. Fig. 4 and Fig. 5 show the mean average precision for both types of classes.

Mean Average Precision(mAP) of a custom trained model

The mean average precision is used to evaluate the object detector whether the trained model is overfitting, underfit or a good fit. Referring to Fig. 4, the result of the trained model is underfitting because the loss is about to reach a point of stability whilst the accuracy is still undesirable. However, in Fig. 5, the trained model is considered as a good fit due to the high accuracy obtained during the point of stability.

The mean average precision for a single class shown in Fig. 4 has a very low confidence score due to a very large data collected in a single class called a weapon. The maximum mAP obtained after training was only 25.19%. The training had to stop halfway because the model was extremely underfitted. Inside this class, it consists of a few types of weapons such as handgun, knife, sword, riffle, arrow and many more. These different types of weapons have different shapes and size, making them harder to recognize during the training process.

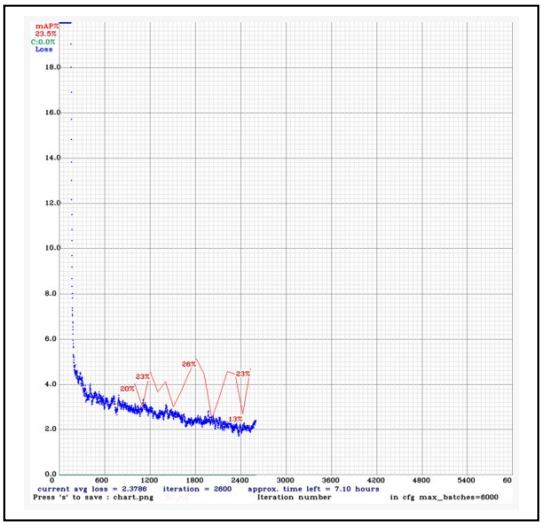

Fig. 4. Single Class mAP

The process flow in Fig. 5 is slightly different than Fig. 4 since there are two types of objects that needs to be detected. It starts by reading the input image or video footage frame by frame. The input is split into grid cells to predict each bounding box simultaneously along with its confidence value. The confidence score indicates how confident the model is when predicting it. Each of the grid cells also predicts the class probability that it belongs to.

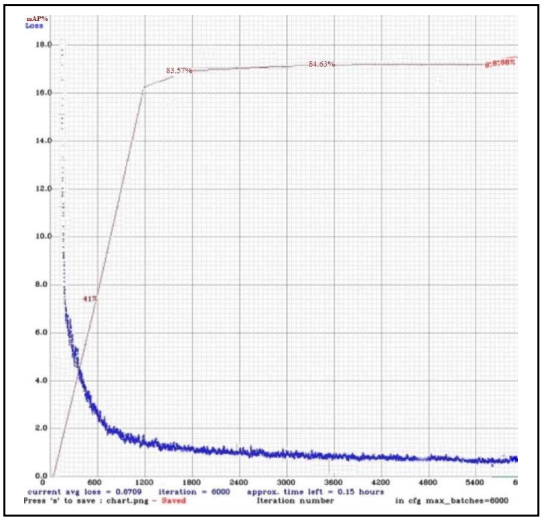

Fig. 5. Multiple Classes mAP.

The predicted class will not be displayed if the confidence score is less than 0.6, similar to single class in Fig. 4. Therefore, each predicted class will be evaluated. If the threshold for a class called handgun is above 0.6, a red bounding box will be displayed. Next, it will check if the threshold for the class called knife is above 0.6, if the result is true, a turquoise bounding box will be displayed. Or else, it will just ignore the predicted bounding box and continue to read the frame.

Object Detection Model comparison

This section compares the result of the detected images using the TensorFlow framework. The obtained darknet weights file has been converted to a TensorFlow file to run this object detection. To compare the effectiveness of both systems, the accuracy is then calculated. Each value of true positive (TP), false positive (FP), true negative (TN) and false negative (FN) in each image or video is counted. The following formulae shown in (1) are used to calculate the accuracy for each system.

$$\text{Accuracy} = (TP+TN)/(TP+TN+FP+FN) \quad (1)$$

TABLE I. ACCURACY OF SINGLE CLASS OBJECT DETECTION

No	Image	TP	FP	TN	FN	Accuracy
1	Self-taken	1	0	5	3	66.67%
2	Handgun and knife	2	1	1	0	75%
3	Handguns	7	0	0	0	100%
4	Knives	7	0	0	2	77.78%
5	Random position 1	3	1	3	2	66.67%
6	Random position 2	2	1	5	1	77.78%

Based on the results shown in Table I, the accuracy fluctuates regularly depending on the types of an image given. Six different types of images are tested to identify the problem. Based on the findings, this is caused by the object detector not being able to detect an image properly due to a very low mAP score.

Fig. 6. Result of object detection on the self-taken image.

Fig. 6 shows the result of object detection when executes on a self-taken image. Knives are placed in

random positions and random items are added to test the effectiveness of the detection. For this single class object detection, only one weapon is detected with a confidence value of 98%.

Fig. 7. Handgun and knife detected in one bounding box along with a torchlight.

In Fig. 7, both weapons are being detected by a single bounding box with a low confidence value of 66%. However, this result is not accurate as there is a non-weapon inside the same red bounding box.

Fig. 8. Image of handguns detected correctly.

Fig. 9. Image of knives detected together in one bounding box but fails to detect all.

All handguns have been correctly detected by the object detection model in Fig. 8 with a high confidence value ranging from 86% to 99%. However, in Fig. 9 most knives are being detected by a single bounding box with a confidence value of 88%. The result is considered inaccurate as two knives are not being detected. Based on the findings, the single class object detection model has difficulty in recognizing precisely when weapons are put closed together unlike in Fig. 8 where the weapons are arranged with a small distance from one another.

Fig. 10. The image of several weapons failed to be detected by the object detection model.

In Fig. 10, only three weapons have been successfully detected by the object detection model. The confidence value ranges from 71% to 86%. However, the other two weapons are failed to be detected by the object detection model.

Fig. 11. Image of a fidget spinner being detected as a weapon.

In Fig. 11, weapons are placed in many different positions with other random items. The object detector failed to detect one knife at the bottom of the image but wrongly detected a fidget spinner as a weapon with a confidence value of 68%. Other than that, the object detector managed to successfully detect the other weapons with a high confidence value.

TABLE II. ACCURACY OF TWO CLASS OBJECT DETECTION

No	Image	TP	FP	TN	FN	Accuracy
1	Self-taken	4	0	5	0	100%
2	Handgun and knife	1	0	2	1	75%
3	Handguns	7	0	0	0	100%
4	Knives	9	0	0	0	100%
5	Random position 1	5	0	4	0	100%
6	Random position 2	2	0	6	1	88.89%

However, the results in table II depicts much higher accuracy in predicting the object. The same images are being tested on this object detection model, but the results produce major difference from the single class object detection model. This is due to the sufficient mAP score attained after the training.

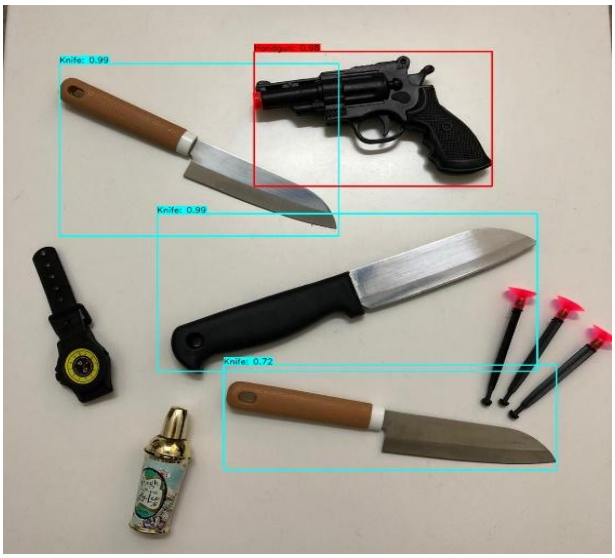

Fig. 12 The self-taken image that has been successfully detected by the object detection model.

The same self-taken image is used in this two class object detection model. The result turns out to be 100% accurate in Fig. 12 even though weapons are placed at random positions. The confidence level is also high for three of the objects but only one is slightly lower than the rest.

Fig. 13 Handgun has been detected but fails to detect the knife.

In Fig. 13, only the handgun was detected with a confidence value of 94%, exhibit an improvement of 28% compared to the previous detection in single class, for this similar image. The knife however could not be recognized by the object detection model due to its position placement.

Fig. 14 Object detection model successfully detected each handgun.

Surprisingly, Fig. 14 and Fig. 8 both obtained a 100% accuracy on this exact image. The findings can be concluded that due to the object being far apart from each other making the object detection model recognize the image even better.


Fig. 15 The object detection model has successfully detected each knife.

In Fig. 15, the multiple class object detection model managed to successfully detect every one of the knives with a confidence level ranging from 65% to 100%. In comparison to Fig. 9, Fig. 15 predicts better and is more accurate as it was able to detect each knife individually.

Fig. 16 The object detection model successfully detects all weapons.

Fig. 17 The object detection model fails to detect an object with a different position.

Fig. 16 managed to detect all of the available weapons in the picture with a very high confidence value as compared to Fig. 10 for the similar image. However, in Fig. 17 the multiple class object detection model failed to detect a knife that is placed vertically. However it managed to detect another two weapons correctly without any false positive when compared to Fig. 11 for the same image in sibgle class detection..

CONCLUSION

To conclude, the objective has successfully achieved as this system can detect two types of weapons specifically, handguns and knives. This approach is beneficial to everyone especially those in the security industry. By comparison, the multiple class object detection model has better mAP scores and result in better detection compared to the single class object detection model. The multiple class object detection model even attained a much higher accuracy compared to the single class object detection model. Overall, the multiple class object detection model has a better performance in detecting objects. However, it can sometimes overlook if the object is moving too fast or the object is placed at a different angle or position.

To improve this system, the multiple class object detection model can be upgraded by making it able to classify more types of weapons. It can also help the model to be more accurate as it will be trained to classify various types of weapons and will make the system more dynamic. Apart from that, this system can be implemented on a microcontroller such as Raspberry Pi for more robust use. A notification system can also be implemented to send notifications once a harmful weapon has been detected. By combining all of these, the AI based surveillance monitoring system can be implemented in places such as banks, shopping malls and even at home.

REFERENCES

[1] [1] K. Yeganegi, D. Moradi, and A. J. Obaid, "Create a wealth of security CCTV cameras Create a wealth of security CCTV cameras," 2020, doi: 10.1088/1742-6596/1530/1/012110.

[2] [2] H. Rama Moorthy, V. Upadhya, V. V. Holla, S. S. Shetty, and V. V. Tantry, "Challenges encountered in building a fast and efficient surveillance system: An overview," Proc. 4th Int. Conf. IoT Soc. Mobile, Anal. Cloud, ISMAC 2020, pp. 731–737, 2020, doi: 10.1109/I-SMAC49090.2020.9243563.

[3] [3] I. Journal and S. Sciences, "Akpauche: International Journal of Arts and Social Sciences, Vo 1, No 2," no. 2, pp. 96–105.

[4] [4] K. Y. Loh and S. C. Reddy, "Understanding and preventing computer vision syndrome," Malaysian Fam. Physician, vol. 3, no. 3, 2008.

[5] [5] R. Hebbalaguppe, "A computer vision based approach for reducing false alarms caused by spiders and cobwebs in surveillance camera networks," 2014.

[6] [6] W. Aitfares, A. Kobbane, and A. Kriouile, "Suspicious behavior detection of people by monitoring camera," Int. Conf. Multimed. Comput. Syst. -Proceedings, vol. 0, pp. 113–117, 2017, doi: 10.1109/ICMCS.2016.7905601.

[7] [7] Y. L. Lai, C. J. Sheu, and Y. F. Lu, "Does the Police-Monitored CCTV Scheme Really Matter on Crime Reduction? A Quasi-Experimental Test in Taiwan's Taipei City," Int. J. Offender Ther. Comp. Criminol., vol. 63, no. 1, pp. 101–134, 2019, doi: 10.1177/0306624X18780101.

[8] [8] E. L. Piza, "The crime prevention effect of CCTV in public places: a propensity score analysis," J. Crime Justice, vol. 41, no. 1, pp. 14–30, 2018, doi: 10.1080/0735648X.2016.1226931.

[9] [9] R. Kakadiya, R. Lemos, S. Mangalan, M. Pillai, and S. Nikam, "AI Based Automatic Robbery/Theft Detection using Smart Surveillance in Banks," Proc. 3rd Int. Conf. Electron. Commun. Aerosp. Technol. ICECA 2019, pp. 201–204, 2019, doi: 10.1109/ICECA.2019.8822186.

[10] [10] C. C. Fung and N. Jerrat, "Neural network based intelligent intruders detection and tracking system using CCTV images," IEEE Reg. 10 Annu. Int. Conf. Proceedings/TENCON, vol. 2, 2000, doi: 10.1109/tencon.2000.888772.

[11] [11] A. Bochkovskiy, C. Y. Wang, and H. Y. M. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection," arXiv, 2020.

[12] [12] "Msia's crime index down significantly, due in part to Sosma, Poca." [Online]. Available: https://www.nst.com.my/news/crime-courts/2020/09/628610/msias-crime-index-down-significantly-due-part-sosma-poca. [Accessed: 08-Jan-2021].

[13] [13] "Open Images Dataset v6 (Bounding Boxes) | Appen Datasets." [Online]. Available: https://appen.com/datasets/open-images-annotated-with-bounding-boxes/. [Accessed: 10-Jan-2021].