Weapon Detection Using Faster R-CNN Inception-V2 for a CCTV Surveillance System

Narit Hnoohom
Image, Information and Intelligence Laboratory, Department of Computer Engineering, Faculty of Engineering, Mahidol University
Nakorn Pathom, Thailand
narit.hno@mahidol.ac.th

Pitchaya Chotivatunyu
Image, Information and Intelligence Laboratory, Department of Computer Engineering, Faculty of Engineering, Mahidol University
Nakorn Pathom, Thailand
pitchaya.cht@student.mahidol.ac.th

Nagorn Maitrichit
Image, Information and Intelligence Laboratory, Department of Computer Engineering, Faculty of Engineering, Mahidol University
Nakorn Pathom, Thailand
nagorn.mat@student.mahidol.ac.th

Virach Sornlertlamvanich
[1]*Asia AI Institute (AAII), Faculty of Data Science, Musashino University, Tokyo, Japan*
[2]*Faculty of Engineering, Thammasat University*
Pathumthani, Thailand
virach@gmail.com

Sakorn Mekruksavanich
Department of Computer Engineering, School of Information and Communication Technology, University of Phayao
Phayao, Thailand
sakorn.me@up.ac.th

Anuchit Jitpattanakul
Intelligent and Nonlinear Dynamic Innovations Research Center, Department of Mathematics, Faculty of Applied Science, King Mongkut's University of Technology North Bangkok
Bangkok, Thailand
anuchit.j@sci.kmutnb.ac.th

***Abstract*— Thailand has faced unrest in recent years, as have other countries around the world. The continuation of present trends means a tendency for an increase in both crimes against people and property. Nowadays, CCTV technology is widely used as surveillance and monitoring tools to help keep people safe. However, most of them still rely primarily on police personnel to inspect the displays. A weapon detection system can reduce the screen-reading workload of police officers with a limited workforce. The integration of weapon detection with CCTV cameras has a role to play in solving the problem. To develop the weapon detection system, the datasets used in this research were collected from 2 public datasets: ARMAS Weapon detection dataset and IMFDB Weapon detection system. The object detection method was used from TensorFlow Object Detection API using 1) SSD MobileNet-V1, 2) EfficientDet-D0 and 3) Faster R-CNN Inception Resnet-V2. For all experimental results, the object detection model is the Faster R-CNN Inception V2 using Dataset 1, ARMAS Weapon detection dataset, with the highest mAP of 0.540 with the Average Precision with 0.5 IoU and 0.75 IoU at 0.793 and 0.627, respectively.**

Keywords—Object detection, Deep learning, CCTV footages, Weapon detection

I. INTRODUCTION

Every year, the number of crimes occurring due to theft or robbery has been increasing gradually. The justice system would benefit from evidence that could be used to arrest and prosecute the offenders. Such evidence could be necessary for the positive identification of offenders or identifying specific crime scene details. There are applications of recording equipment such as Closed-Circuit Television Systems (CCTV) that can be used as a tool for security. This is possible in the form of surveillance through a large number of cameras that record events. A significant advantage of imaging is its reliability and credibility. These variables, however, are frequently dependent on image quality, which must be able to record clear footage in order to unmistakably identify offenders or details such as faces, license plates, or other indications that appear in images. These details of incidents could be used as valid evidence.

CCTV cameras are an important part of solving security problems and are considered one of the most important security requirements [1]. CCTV surveillance systems have become increasingly popular over the last decade. They have been developed as tools in public administration-related activities to maintain public order, such as control and prevention, as well as criminal investigation, including surveillance to regulate unpleasant and anti-social behavior in major cities across the world. In the UK, about 2 million CCTV cameras have been deployed, with 400,000 CCTVs utilized for surveillance.

Thailand is dealing with a variety of criminal offenses, including crimes against life and property, as well as unpleasant behavior that cause dread among individuals. According to the Royal Thai Police's statistics concerning the occurrence of crimes reported in 2016 [2], there were 14,459 crimes and criminal offenses reported across the country. Police officers apprehended 13,645 offenders, or 94.37 percent, and sentenced 5.63 percent of those arrested. There were 2,119 crimes and criminal acts in Bangkok, which is the economic and social capital of Thailand. The police arrested 1,897 offenders, accounting for 89.52 percent of all offenders, while almost 10.48 percent were not apprehended. Furthermore, a survey of Bangkok residents' feelings towards crime indicated that 52.20 percent felt insecure [3]. There is also the case of intentional murder in Nakhon Ratchasima Province [4], which is a crucial case that caused tragedy to the people in the country. The criminal drove a car loaded with military-grade weapons and indiscriminately shot innocent people. The criminal siege of Terminal 21 in Korat resulted in 29 people killed and 58 injured, which was a significant loss.

978-1-6654-1197-4/21/$31.00 ©2021 IEEE

It is imperative to enable technology to assist in surveillance to maintain the safety of the population. At present, CCTV systems have been installed in many provinces including Bangkok, Phuket, etc. Still, there are limitations in some situations, namely the continued reliance on humans to monitor the displays, which is the reason why there may still be shortcomings in some events or cases which are caused by limitations in human capabilities, such as incomprehensible monitoring. Sometimes, an officer cannot collect evidence to prosecute the culprit or offender due to significant groups of people and displays, or because he/she is unable to act when an incident occurs. To improve the efficiency of a CCTV system, it is necessary to use artificial intelligence technology to help detect suspicious behavior, such as checking for people with guns, tracking suspects' movements, and checking for escape routes to help find offenders.

The integration of weapon detection with CCTV cameras has a role to play in solving the problem. A suspect's behavior can be revealed by the weapon detection system, resulting in a decrease in the risk of harm to people who aren't frightened of the law harming people's lives and property, as well as providing protection against potential losses. When surveillance cameras are installed in critical areas of the city, a suspect can be found and the system can notify the police station close to the coordinates of the suspect. This is possible by using CCTV cameras that have recorded the installation point to identify the area in question.

II. Related Work

A weapon detector is an algorithm that detects armed people who are committing or are about to commit a crime. The weapon detector is intended to leverage the existing infrastructure of CCTVs, both public and private, to assist in alerting local security and determining possible criminals. We think that by using the gun detector, we will be able to minimize the number of people needed for CCTV coverage while also improving the functioning of current CCTVs.

A. Image Processing for Object Detection

Image processing has been used for various field of vision tasks. For weapons detection, image processing and machine learning methods are widely used to detect weapons by the shapes of the objects. Tiwari et al. [5] proposed a scheme for the automatic detection of guns using a hybrid approach of color-based segmentation and interest point detector. The authors have used a combination of Harris interest point detectors and Fast Retina Keypoint (FREAK) descriptors to detect interest points and extract the features used for matching with gun descriptors. Color-based segmentation is performed using a k-means clustering algorithm to eliminate unrelated colors or objects present in the image. Interest point features of object boundary are then extracted, which are used to match a stored descriptor to find any similarity with the gun. If the similarity score is greater than 50 percent, then the system will raise an alarm signal.

While Speeded up robust features (SURF) is popularly used to detect objects within the image, SURF interest point detector using a blob or edges to locate the object (gun) in the segmented images. Tiwari et al. [6] presented a framework that uses color-based segmentation to remove unrelated objects from a picture using the K-mean clustering technique utilizing SURF and constructed and evaluated the system on self-collected sample images of firearms. Further, the system performs very well under the method and improves significantly when images have a variety of characteristics. Matching is committed in order to compute the similarity score between the stored descriptors of gun and blob. The SURF features of an object's boundary are used for matching because the inner texture of the object body varies from object to objects of the same kind, leading to different SURF features for the same interest point different images, although the outline of the object remains constant or varies slightly.

B. Object Detection using Deep Learning

Deep learning has recently become popular for artificial intelligence problems. Deep learning is a method that trains a model to comprehend data characteristics without the requirement for self-selected features during training. The algorithm can understand data that is fed through the network. Deep learning has given rise to several exciting developments in the computer vision sector, such as self-driving automobiles [7] and alpha-go [8], which can compete against world go champions. Carrobles et al. [9] proposed gun and knife detection based on Faster Region-based Convolutional Neural Network (Faster R-CNN) for video surveillance to boost the accuracy and usability of CCTV cameras for deep learning in a weapon detection study. This study provides a gun and knife detection system based on the Faster R-CNN technology. Two techniques were examined using GoogleNet and SqueezeNet architectures as convolutional neural network (CNN) bases.

Bhatti et al. [10] suggested a deep learning-based weapon detection system for real-time CCTV recordings with the goal of recognizing a pistol or comparable fired portable weapon in real-time. As a detection algorithm, the region proposal is applied. Under the region proposal approach, Single Shot MultiBox Detector (SSD) Mobilnet-V1, You only look once (Yolo) V3, Faster R-CNN-Inception ResNet-V2, and Yolo-V4 are implemented for real-time detection. Three separate datasets are used to train the algorithms. Each one has a different set of data and a specific train-test dataset ratio. The Yolo-V4 outperformed another model in the object detection model evaluation. González et al. [11] introduced real-time gun detection in CCTV: An open problem for enhancing small object recognition and increasing data. The authors created synthetic datasets by generating virtual environments with the Unity game engine. They also included data from a simulated attack on the University of Seville, which was recorded and gathered. The research employs a Faster R-CNN architecture based on the Feature Pyramid Network (FPN) for usage in real-time CCTV. For boosting the precision of recognizing small items from experiments, a combination of first training synthetic images and then real photos is the best approach. However, several tests struggle to recognize distant objects because the dataset is made mostly of foreground photos. Because there is no class for classifying other items, it can lead to false-positives.

III. Object Detection Training

A. System Overview

For the overview of the system, the crime intention detector system is an artificial intelligence platform for detecting criminal intentions and events. As the CCTV footage is recorded via video recorder, each frame from the video is used to train the artificial intelligence to detect crime events as shown in Figure 1.

The system uses weapon detection to detect weapons on the image. The system will analyze the results from each

process and conclude the result of the person in the event has a probability to commit a crime.

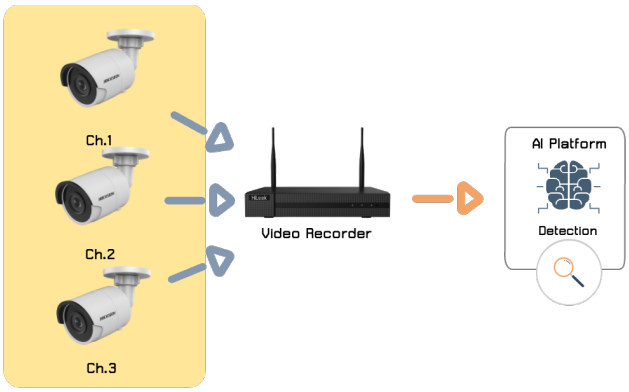

Fig. 1. Process of the weapon detection system

B. Data Collection

In this research, data collection for crime intention was collected on images of crime intended using guns. The authors collected data for armed crime intention in public using a public image dataset from the Weapon detection dataset [13] and Weapon detection system [14].

1) Dataset 1: ARMAS Weapon detection dataset

The dataset comprises a total of 3,000 pistol images. The images contained in this dataset comprise pistols which are clearly visible. In some images, the pistols are held by a person, which can still be visualized clearly. Examples are illustrated in Figure 2.

Fig. 2. Public Dataset Example

2) Dataset 2: IMFDB Weapon detection system

The dataset is composed of pistol images and rifle images from the Internet Movie Firearms Database (IMFDB) dataset, and some CCTV footage comprised of 4,940 images. The images have various sizes from 99 × 93 pixels to 6,016 × 4,016 pixels with different ranges of vision and composition. Examples are illustrated in Figure 3.

Fig. 3. Public Dataset Example

After collecting datasets, the authors preprocessed the images in datasets to suit the object detection training methods, then partitioned the dataset into a training set and testing set using the ratio of 80:20 (80 percent partitioned to a training set and 20 percent used as a testing set). Each dataset was partitioned as mentioned, with the number of each training set and testing set as shown in Table I.

TABLE I. WEAPON DETECTION DATASET PARTITION

Dataset	Total	Train	Test
Dataset 1	3,000	2,400	600
Dataset 2	4,940	3,952	988

C. Weapon Detection Model Training

Weapon detection model training needs informative data to learn the images. The dataset must be prepared correctly. The dataset is composed of the image files and label files which indicate the weapon object in the images, which are then partitioned into the training set and testing set. The partitioned dataset is needed to convert into a TFRecord file, which is the data file format used for TensorFlow Model Zoo object detection model training.

This research uses the pre-trained model from the TensorFlow Model Zoo featuring:

1) *SSD MobileNet-V1*
2) *EfficientDet-D0*
3) *Faster R-CNN Inception Resnet-V2*

In the model training process, the divided TFRecord files were used for feeding image information through the network. The configuration pipeline file is used for tuning the iteration and other values for model training. The label map file is also needed for indicating the type of objects for the model learning. For evaluating the weapon detection model, the results value will indicate the location of the detected weapon on images and classified object type. To analyze the model correctness, the Mean Average Precision (mAP) and the Intersection over Union (IoU) are involved.

D. Model Evaluation

The object detection result returns the predicted class and bounding box of the object. The evaluation is concerned with the standard loss function for classification of the prediction class and localization of the bounding box from the

TensorFlow pre-trained Object Detection Application Programming Interface (API) [12].

IV. Results

A. Model Evaluation

From the weapon detection experiment, the public datasets were used from 2 sources composed of the Weapon detection dataset [13] and the Weapon detection system [14]. The pre-trained model from the TensorFlow framework was used for detection training:

1) SSD MobileNet-V1

2) EfficientDet-D0

3) Faster R-CNN Inception Resnet-V2

The model will adjust the localization loss and classification loss to improve the correctness of the data understanding of the detection model. The mentioned losses can indicate the mathematical efficiency of incorrect prediction from classifying types. From Figure 4, the X-axis indicates the model training step, while the Y-axis indicates the classification loss. The orange plotted line shows the training loss from MobileNet-V1, the blue plotted line shows the training loss from Faster R-CNN Inception-V2, while the red plotted line shows the training loss from EfficientDet-D0.

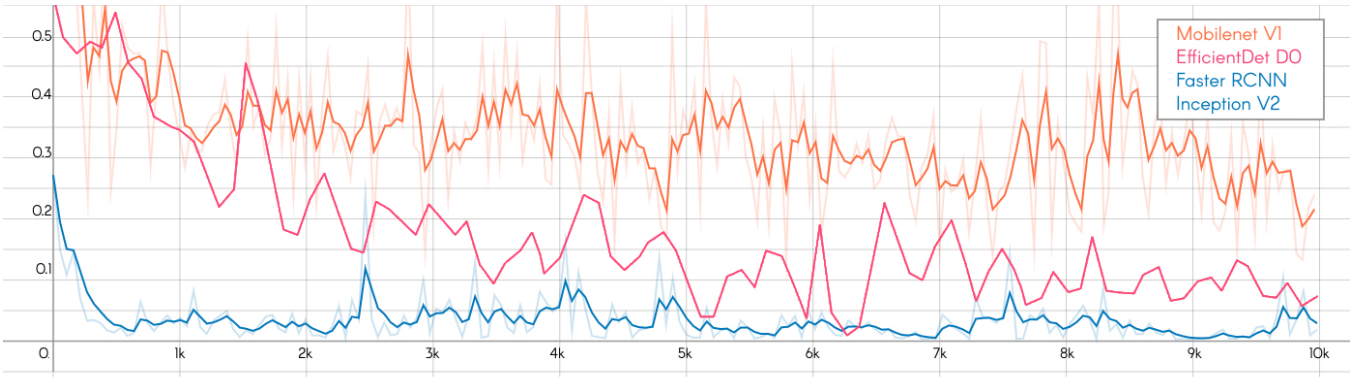

Fig. 4. Classification loss

From Table II, the weapon detection models were trained with Dataset 1 comprising 3,000 pistol images, 2,400 images for training, and 600 images for testing. The models achieved the highest mAP of 0.562, mAP at 0.5 IoU of 0.787, and mAP at 0.75 IoU of 0.618 from SSD MobileNet-V1.

TABLE II. Weapon Detection Model Evaluation of Dataset 1

Architecture	mAP	0.5 IoU	0.75 IoU
SSD MobileNet-V1	0.562	0.787	0.618
EfficientDet-D0	0.431	0.771	0.433
Faster R-CNN Inception-V2	0.540	0.793	0.627

Table III shows the corresponding average precision (AP) values per pistol class for the evaluation of the mean average precision value from training Dataset 1.

TABLE III. Weapon Detection Model Evaluation of Dataset 1

Architecture	Precision	Recall	F1-score
SSD MobileNet-V1	78.7%	69.5%	73.81%
EfficientDet-D0	77.1%	59.9%	67.42%
Faster R-CNN Inception-V2	79.3%	68.6%	73.56%

From Table IV, the weapon detection models were trained with Dataset 2 comprising 4,990 images, 3,952 images for training, and 988 images for testing. The models achieved the highest mAP of 0.456, mAP at 0.5 IoU of 0.5526, and mAP at 0.75 IoU of 0.4591 from EfficientDet-D0.

TABLE IV. Weapon Detection Model Evaluation of Dataset 2

Architecture	mAP	0.5 IoU	0.75 IoU
SSD MobileNet-V1	0.380	0.669	0.342
EfficientDet-D0	0.456	0.5526	0.4591
Faster R-CNN Inception-V2	0.335	0.643	0.316

Table V shows the corresponding AP values for the pistol class for the evaluation of the mAP value from training Dataset 2.

TABLE V. Weapon Detection Model Evaluation of Dataset 2

Architecture	Precision	Recall	F1-score
SSD MobileNet-V1	66.9%	56.2%	61.08%
EfficientDet-D0	55.3%	59.9%	57.51%
Faster R-CNN Inception-V2	64.3%	53.1%	58.17%

From the evaluation of the weapon detection model, the MobileNet-V1 provided higher detection precision than in EfficientDet-D0 and Faster R-CNN Inception Resnet-V2 for detecting gun images. From Figure 5, the EfficientDet D0 is unable to detect the labels on gun images, while Faster R-CNN

Inception Resnet-V2 shows false positive detection on non-pistol objects.

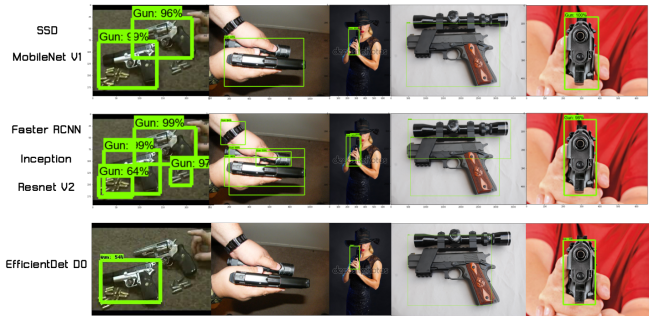

Fig. 5. Weapon detection model on example images

Because each study has a unique dataset, models, and measures for evaluating performance, the results may not be comparable. Each research attempt has its own set of testing circumstances, such as focusing just on pictures, videos, or high-resolution photographs. The performance metric implemented in several researches is precision, while others use precision or mean average precision (mAP). However, mAP is the most widely used. Thus, we compared the findings in terms of mAP and precision at a standard IoU threshold of 0.50 as shown in Table VI.

TABLE VI. MODEL EVALUATION

Study	Algorithm	Precision	mAP
Carrols et al. [5]	Faster R-CNN GoogleNet	55.45%	-
Bhatti et al. [10]	Faster R-CNN Resnet50 FPN (Handgun dataset)	88.1%	0.652
	Faster R-CNN Resnet50 FPN (Mock Attack Cam)	3.4%	0.009
González et al. [11]	SSD MobileNet-V1	62.79%	-
	Faster R-CNN Inception-V2	86.38%	-
Proposed method	SSD MobileNet-V1 (Dataset 1)	78.7%	0.562
	EfficientDet-D0 (Dataset 1)	77.1%	0.431
	Faster R-CNN Inception-V2 (Dataset 1)	79.3%	0.540

Compared to Carrobles et al. [5], gun and knife detection based on Faster R-CNN for video surveillance uses the CCTV viewing angle images along with the image augmentation method for preprocessing. From the results, our methods using Faster R-CNN Inception Resnet-V2 performed precision at 0.5 IoU of 79.3 percent, better than using Faster R-CNN GoogleNet, which acquired a result of 55.25 percent. Bhatti et al. [10] used Faster R-CNN Resnet50 FPN training on different datasets; the handgun dataset achieved precision at 0.5 IoU of 88.1 percent and mAP at 0.652, while our dataset with a similar structure of images dataset performed precision at 0.5 IoU of 78.7.1 percent and mAP at 0.562. González et al. [11] achieved precision at 0.5 IoU of 62.79 percent from SSD MobileNet-V1 training for weapon and non-weapon datasets. Compared to our method, SSD MobileNet-V1 achieves 78.7 percent precision at 0.5 IoU.

V. CONCLUSION AND DISCUSSION

A. Conclusion

This research presents a weapon detection model using object detection. The object detection method is based on Model Zoo Object Detection API for specifying weapons in an image. The model provided precise prediction and localization of weapon type when using Dataset 1 (ARMAS dataset), which involved a medium to large object. Finally, the experimental report indicated the potential of deep learning algorithms for resolving crime-intended events.

B. Recommendations and Future Developments

The experiments ended as the result of object detection on a pistol using a different characteristic of a dataset. The images used in dataset 2 were varied in size, though most were small in size, as provided from the public datasets. When training object detection using a bigger input size than the input image resolution, the object detection is mainly focused on small object detection and becomes more challenging to localize and classify correctly. To improve the efficiency of the model, the dataset should be preprocessed before training or use self-acquired CCTV images with default resolution.

ACKNOWLEDGMENT

The authors gratefully acknowledge the financial support provided by the Thammasat University Research fund under the TSRI, Contract No. TUFF19/2564, for the project of "AI Ready City Networking in RUN", based on the RUN Digital Cluster collaboration scheme, and supported by the Department of Computer Engineering, Faculty of Engineering, Mahidol University.

REFERENCES

[1] Deisman, Wade. "cctv: Literature review and bibliography." In Research and Evaluation Branch, Community, Contract and Aboriginal Policing Services Directorate. Ottawa: Royal Canadian *Mounted*. 2003.

[2] National Statistical Office, Statistics of reports and arrests of violent and shocking cases classified by the type of cases reported in each province, 2007 - 2016, Available: http://statbbi.nso.go.th/staticreport/page/sector/th/09.aspx [2020, Febuary 25]

[3] Analysis and Evaluation Group, A survey of people's feelings of terror about crime in Bangkok. 2014 Annual Research Report, Police Strategic Research Division, 2557.

[4] Thairath Online, The government pays the victims of the mad sergeant millions each. The other 3 bodies are not given. Available: https://www.thairath.co.th/news/local/northeast/1771045 [2020, Febuary 25]

[5] Tiwari, R. K., Verma, G. K., "A Computer Vision based Framework for Visual Gun Detection using SURF", 2015. Available: https://www.researchgate.net/publication/281776103_A_computer_vision_based_framework_for_visual_gun_detection_using_SURF

[6] Tiwari, R. K., Verma, G. K., "A Computer Vision based Framework for Visual Gun Detection Using Harris Interest Point Detector", 2015. Available: https://www.sciencedirect.com/science/article/pii/S1877050915014076

[7] Huval, B., & et, a. (2015). An empirical evaluation of deep learning on highway driving. arXiv preprint. Available: https://arxiv.org/abs/1504.01716.

[8] Silver, D., & et, a. (2016). Mastering the game of Go with deep neural networks and tree search. Nature, 484-489.

[9] M. Milagro Fernandez-Carrobles, Oscar Deniz and Fernando Maroto "Gun and knife detection based on Faster R-CNN for video surveillance", 2019. Available: https://www.researchgate.net/publication/335969239_Gun_and_Knife_Detection_Based_on_Faster_R-CNN_for_Video_Surveillance

[10] Navalgund, U. V., Priyadharshini, K. "Crime Intention Detection System Using Deep Learning", 2018. Available: https://ieeexplore.ieee.org/document/8821168

[11] Gonzáleza, J. L. S., Zaccaroa, C., Álvarez-García. J. A., Morilloa, L. M. S., Caparrini, F. S., "Real-time gun detection in CCTV: An open problem", 2020. Available: https://www.sciencedirect.com/science/article/abs/pii/S0893608020303361

[12] "tensorflow/models," *GitHub*. https://github.com/tensorflow/models (accessed Mar. 09, 2020).

[13] Abhishek Annamraju, "Weapon Detection Dataset", 2019. Available: https://www.kaggle.com/abhishek4273/gun-detection-dataset [2021, March 20]

[14] Samuel Mohebban, "Weapon Detection System", 2020. Available: https://github.com/HeeebsInc/WeaponDetection [2021, March 20]