International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 09 Issue: 07 | July 2022 www.irjet.net p-ISSN: 2395-0072

Weapon Detection and Classification in CCTV Footage

Tejashwini PS[1], Saatvik Girish Nargund[2], Sumuk K[2], Rohith G R[2], Trinetra BM[2]

[1]*Assistant Professor, Department of CS&E, BIT, Bangalore, Karnataka, India*
[2]*Student, Department of CS&E, BIT, Bangalore, Karnataka, India*

---***---

Abstract - *Closed Circuit Television (CCTV) cameras are being used for surveillance and to monitor activities i.e. robberies, but these cameras still require human supervision and intervention. We need a system that can automatically detect these illegal activities. This work focuses on providing a secure place using CCTV footage as a source to detect harmful weapons by applying the state of the art open-source deep learning algorithms. We have implemented a classification model with different classes of weapons and relevant confusion objects inclusion concept is introduced to reduce false positives and false negatives*

Key Words: Machine Learning, Convolutional Neural Network, Weapon Detection, Pistol Detection, Knife Detection

1. INTRODUCTION

Security and safety is a big concern for today's modern world. For a country to be economically strong, it must ensure a safe and secure environment for investors and tourists. Having said that, Closed Circuit Television (CCTV) cameras are being used for surveillance and to monitor activities i.e. robberies. It is considered as one of the most important evidence in law enforcement agencies and courts. Therefore, the amount of CCTV cameras installed have increased in number all over the world. This has made the public feel much safer, and decreased crime in many areas all over the world. As a result of the increase in number of CCTV cameras, the number of screens that have to be monitored by a single CCTV operator has increased a lot. One operator cannot be expected to monitor many screens at a single time. Also, he cannot constantly monitor the footage throughout the day Moreover, it is difficult and expensive to monitor hundreds, or even thousands of CCTV video footage in an area. Therefore, there is an increasing demand to automate CCTV surveillance.

1.1 Objectives of the system

The main aim of automated CCTV surveillance is to alert the CCTV operator whenever there is a dangerous situation. A dangerous situation refers to a person or a group of people attacking, creating fear or disturbances with weapons like knives and guns. These kinds of situations can be detected by automated systems. The systems work by using object detection algorithms to classify objects detected in the video footage. Our proposed system will detect and classify any weapons which may be present in the CCTV footage. Weapons can include knife and pistols. Whenever a weapon is detected, it will alert the user (CCTV operator). It will also be possible to add exceptions to the rules if the user wishes.

2. RESEARCH ELABORATION

We have researched various papers to get to understand the different methods used by various authors in their own weapon detection systems. Aarchi Jain et. al. explored the usage of gun detection using the Haar Cascade Classifier [1]. The low accuracy on pistols was a drawback in this system. Harsh Jain et. al. implemented a system using a CNN based SSD algorithm, which offered improved precision [2]. Pawel Donath et. al. presented a neural network based system using a custom neural net inspired by AlexNet and VGGNet architectures [3]. Michal Grega et. al. have used descriptors like edge histogram and homogenous texture, along with SVM, in their system to detect pistols and knvies [4]. JLS Gonzalez, in their proposed system, have used faster RCNN, along with creation of synthetic datasets using Unity game engine to compensate for the lack of good quality date.

Muhammad Tahir Bhatti et. al. implemented and tested various models for Pistol Detection in CCTV footage [6]. All the models tested were based on convolutional neural networks (CNN). Kushagra Yadav et. al. used N stage learning, along with ADADELTA technique to train a neural network to detect different types of weapons and knives [7]. Jacob Rose et. al. propsed a Faster RCNN based model using the ResNet – 50 base network for weapon detection [8]. This model was trained on COCO dataset. Mitchell Singleton utilized the MobileNetV1 Neural Network to identify handguns in in various orientations, shapes, and sizes [9]. Jesus Ruiz-Santaquiteria et. al. combined, in a single architecture, both weapon appearance and human pose information to better detect weapons in an image [10].

3. METHODOLOGY

The techniques used in this paper mainly center around object recognition. It is a complex task which actually involves 3 sub tasks: Image Classification, Object Localization, and Object Detection. Image Classification predicts the class of an object in an image. Object Localization locates the presence of objects in an image and indicate their location with a bounding box. Object Detection locates the presence of objects with a bounding box and detect the classes of the located objects in these boxes.

One of the first object detection algorithms was AlexNet. It won the ImageNet competition. This was the beginning of object detection in deep learning. Initially, Sliding window models were used. Since it is an exhaustive search over the whole picture for objects, it has high memory and CPU costs, and is inefficient. The next stage of algorithms to be used were region proposal algorithms. They used selective search algorithms. However, their performance was not suitable for real time systems.

One of the fastest series of algorithims today is YOLO family. YOLO ("You Only Look Once") is an effective real-time object recognition algorithm. It gives a much better performance on all the parameters we discussed along with a high fps for real-time usage. YOLO algorithm is an algorithm based on regression. Instead of selecting the interesting part of an Image, it predicts classes and bounding boxes for the whole image in one run of the Algorithm. YOLO doesn't search for interested regions in the input image that could contain an object, instead it splits the image into cells, typically a 19x19 grid. Each cell is then responsible for predicting K bounding boxes. An Object is considered to lie in a specific cell only if the center co-ordinates of the anchor box lie in that cell. Due to this property the center co-ordinates are always calculated relative to the cell whereas the height and width are calculated relative to the whole Image size. During the one pass of forwards propagation, YOLO determines the probability that the cell contains a certain class. The class with the maximum probability is chosen and assigned to that particular grid cell. Similar process happens for all the grid cells present in the image. The latest version of YOLO, YOLOv5 has been used in our system.

3.1 Training Mechanism

The training mechanism used in a machine learning project, such as ours, is given in Fig 1. It starts with defining a problem, finding the required dataset, applying pre-processing methods, and then finally training and evaluating the dataset. If the evaluation is. correct then we save those weights as a classifier. If those weights are incorrect, then the parameters are tuned, and the whole process is repeated. Some of the parameters we have experimented with on this project are: Weight Decay, Box Loss Gain, etc. The trained weights are evaluated on real life CCTV footage gathered from various sources like Youtube, etc.

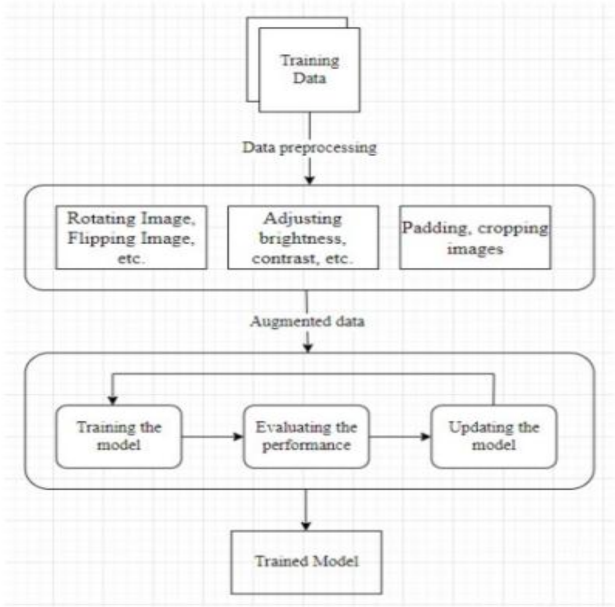

Fig -1: Training Mechanism

3.2 Dataset Construction

We have collected data from multiple datasets so that our model can be trained on the maximum amount of data for the best performance and accuracy. The dataset has been collected from different places around the internet, such as: extracted from YouTube CCTV videos, through GitHub repositories, data by the University of Granada research group, and internet movie firearm database imfdb.org. In all these, only the dataset which contained pistols and knives were chosen. We have also included confusion objects in the dataset, like phones, wallets, etc. This is to include the category of objects which may be confused with knife and pistol, so that the model does not confuse them for it. The split is given in Table 1.

Table -1: Dataset used for training

Category	Total Images	Training Images	Test Images	Split Ratio
Pistol	2971	2526	445	15%
Knife	2178	1851	327	15%

To increase the size of the dataset, data preprocessing was done. The pre-processing process involves data cleaning, standardization, processing, extraction and choice of features, etc. The final training dataset is the result of pre-processing processes applied to the collected dataset. It is necessary for better training of a model, so multiple steps must be taken to augment the size of the dataset. Some of the data augmentation techniques that we have used are: image rotation, image cropping, image scaling. After creating the augmented dataset, we fed it into the YOLOv5 model with

International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 09 Issue: 07 | July 2022 www.irjet.net p-ISSN: 2395-0072

default hyperparameters. After a few iterations of training and testing, we finally obtained good results for the pistol and knife detection model.

4. METHODOLOGY

After obtaining the model, it was tested on a variety of real life CCTV footage containing knives and guns. These footages are of usually low resolution and low quality, so it allowed us to see the speed and accuracy of the model. The confusion matrix is given in Fig 2.

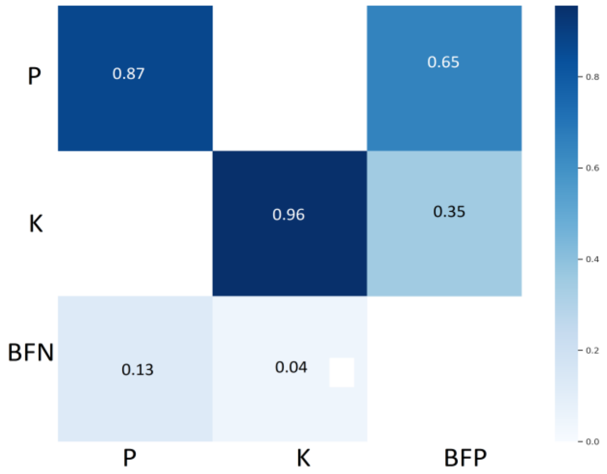

Fig -2: Confusion Matrix

The confusion matrix for the dataset is given in Figure 2. For the pistol class, 87% of pistols which were there in the picture were predicted correctly. For the knife class, 96% of knives which there in the picture were predicted correctly. BFN and BFP represent background false negatives and background false negatives, respectively. Though the algorithm was able to predict almost all the instances in which there were pistols or knives, in some instances it predicted knives and pistols where there were not any. It indicates that further improvement is required to reduce the number of false positives. Moreover, the confidence score might need to be increased if the system is actually used in practice to reduce the number of false positives.

The other results are given in a graph form in Figure 3. The first graph is the precision vs. epochs. The second graph is the recall vs. epochs of the model. The mean average precision at 0.5 and at 0.95 are also given. mAP stands for mean average precision. Average precision (AP) computes the average precision value for all recall values from 0 – 1. It is basically the area under the precision recall curve. The mean average precision is the mean of the APs for all possible classes. In this case, it is the average of APs for pistol and knife classes. The best results for the model are given in Table 2

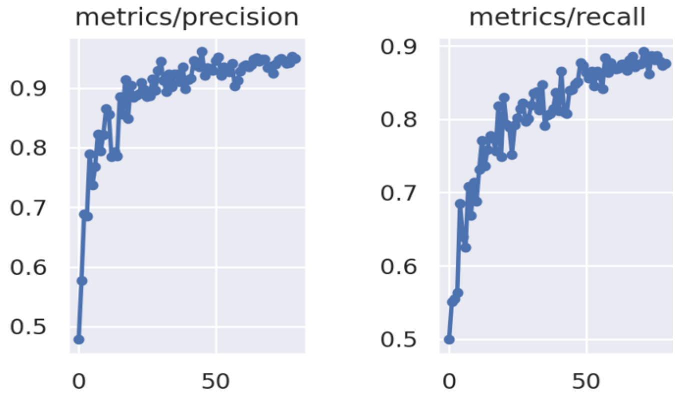

Fig -3: Precision and Recall vs. Epoch Graphs

Table -2: Best results in each category

Precision	Recall	mAP@0.5	mAP@0.95
0.95353	0.88723	0.94576	0.61443

5. CONCLUSION

Weapon detection plays a vital role in safety, security, and surveillance management. In this advanced time of observation and security the quantity of Closed-Circuit Television (CCTV) conveyed out in the open and private places, have expanded exponentially and this has led to a challenge to humans to spot weapon in so many cameras. This paper presents a real-time framework and method that is designed for security purpose. This method was built to identify and classify knives and pistols in a video, as they are the most commonly used weapon types in attacks. It reaches an accuracy of 87% on pistols, and 96% on identifying knives. The salient features in this model are that the weapon can come at any angle but the model can still detect it. Though it has a problem with false positives, with more training and experimentation, it can surely be improved.

ACKNOWLEDGEMENT

We would like to convey our sincere gratitude to our college the Bangalore Institute of Technology, and Dr. Aswath M U, Principal, Bangalore Institute of Technology for being kind enough to provide an opportunity and platform to complete this project. We would also like to thank Dr. Girija J, Professor and Head of the Department for Computer Science and Engineering, Bangalore Institute of Technology for her constant encouragement and making us believe in ourselves. We are most humbled to mention the enthusiastic influence provided during the development and ideation phase by our guide Prof. Tejashwini P S, Assistant Professor, Department of Computer Science and Engineering for her ideas, time to time suggestions and co-operation shown during the venture and helping make this project a success.

REFERENCES

[1] Jain A., Aishwarya and G. Garg, "Gun Detection with Model and Type Recognition using Haar Cascade Classifier," in 2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT), pp. 419-423, 2020.

[2] H. Jain, A. Vikram, Mohana, A. Kashyap and A. Jain, "Weapon Detection using Artificial Intelligence and Deep Learning for Security Applications," in 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), pp. 193-198, 2020.

[3] Donath P., Grega M., Guzik P., Król J., Matiolański A., Rusek K., "Dangerous Tool Detection for CCTV Systems," in Multimedia Communications, Services and Security. MCSS 2020. Communications in Computer and Information Science, vol 1284. Springer, Cham, 2020.

[4] Grega, M.; Matiolański, A.; Guzik, P.; Leszczuk, M. "Automated Detection of Firearms and Knives in a CCTV Image," in Sensors. 16. 47. 2016

[5] Jose L. Salazar, González, Carlos Zaccaro, Juan A. Álvarez-García, Luis M. Soria Morillo, Fernando Sancho Caparrini, "Real-time gun detection in CCTV: An open problem," in Neural Networks, Volume 132, pp. 297-308, 2020

[6] M. T. Bhatti, M. G. Khan, M. Aslam and M. J. Fiaz, "Weapon Detection in Real-Time CCTV Videos Using Deep Learning," in IEEE Access, vol. 9, pp. 34366-34382, 2021.

[7] K. Yadav, R. Punia and N. Raghava, "Towards the Law Enforcement of Firearms," in 2020 IEEE 17th India Council International Conference (INDICON), pp. 1-6, 2020

[8] J. Rose, T. Bourlai and J. A. Loudermilk, "Assessment of Data Augmentation Techniques for Firearm Detection in Surveillance Videos," in 2020 IEEE International Conference on Big Data, 2020, pp. 1838-1846

[9] Singleton M., Taylor B., Taylor J., Liu Q, "Gun Identification Using Tensorflow," in Machine Learning and Intelligent Communications. MLICOM 2018

[10] J. Ruiz-Santaquiteria, A. Velasco-Mata, N. Vallez, G. Bueno, J. A. Álvarez-García and O. Deniz, "Handgun Detection Using Combined Human Pose and Weapon Appearance," in IEEE Access, vol. 9, pp. 123815-123826, 2021