Grenze International Journal of Engineering and Technology, June Issue

Weapon Detection in Surveillance System using Deep Learning

Thota Sai Lalith Prasad[1], Narsimha[2] and B Venkata Srinivasulu[3]
[1]Assistant Professor-Artificial Intelligence & Data Science Department, Vignan Institute of Technology and Sciences, Telangana, India.
Email: lalithresearch15@gmail.com
[2]PG Scholar-Artificial Intelligence & Data Science Department, Vignan Institute of Technology and Sciences, Telangana, India
Email: vogumala2000@gmail.com
[3]Associate Professor-Vignan Institute of Technology and Sciences, Telangana, India.
Email: bysanisrinu13@gmail.com

***Abstract*— Software capable of detecting knives, guns, and other weapons is a personal blessing in order to ensure that the general population is safe. In this paper, the author illustrates the idea of our surveillance system weapon detection system, which is grounded in the deep learning concept. It employs the CNN object recognition algorithms and the You Only Look Once (YOLO) with a view of identifying and analysing possible dangers. The development of safe space is our first priority which can be achieved by use of CCTV footage to identify possible dangerous weapons because this is a badly needed service in contemporary world. Provided that the real-time object detection technology would be able to enhance the surveillance, then YOLO can become a game-changer.**

***Index Terms*— Weapon Detection, Surveillance Systems, Deep Learning, Object Detection, YOLO, CNN, Public Safety.**

I. INTRODUCTION

All using the deep learning algorithms, security applications are able to locate weapons [1]. Computer software based on visuals are able to unravel the action in surveillance video [2]. During the initial stage of video surveillance, guards needed to have a quick glimpse at a film to identify any weapon influence and react to it [3]. The idea to use the deep learning algorithms to the data obtained by real-in-time and implemented security cameras to support the automatization and enhancement of weapon recognition is brilliant [4]. By doing so, we will be able to avoid the problems of the human analysis [5]. The current developments in technology have seen such security cameras installed in various environments with ease and this can reduce the chances of criminals [6]. The security personnel require that the strategically placed surveillance equipment be monitored constantly [7]. Security personnel are supposed to search footage, analyze it and gather evidence on the scene of a crime after an event [8]. Someone must have someone to be in charge and establish a structure at the crime scene [9]. When the security personnel are notified promptly on the presence of dangerous objects through the program, they can be able to intervene in time to ensure that the criminal does not execute the crime [10]. That is why it is essential to establish the system that may help detect the items that might be harmful.

Grenze ID: 01.GIJET.12.2.331_22

It is common consensus that deep learning has significantly improved the efficiency of the security control systems. Deep learning is one of the most glaring fields of machine learning. In case of deep learning, feature extraction, and feature modification, it makes use of multi-layer architecture of non-linear processing units. The data may be accessed in various feature layers due to its nature of deep learning. Convolutional neural networks (CNNs) represent a type of deep learning that becomes trained through exploring the most basic representations of data. To term an image as a feature a feature cluster, a set of custom shapes, or a set of values of pixel density, we are working with the wrong data structure. In the recent past, deep learning has been the trend in various aspects including object identification, image segmentation and classification. YOLO Results using the techniques in traditional image processing tasks such as detection, classification, and segmentation give promising results.

II. Literature Survey

Ref. [11], an item identification system based on garment disguise was displayed live. The YOLO technique of passive millimetre wave photography used a limited dataset and it entailed the placing of metallic weapons on human bones [12]. Second, the Single MultiBox Detector method is evaluated with the help of the PMMW dataset that includes YOLOv3-53, SSD-VGG16, and YOLOv3-13. The weapon recognition accuracy was personalized to 95 wit an average accuracy of 36 frames per second.

Ref. [13] report the results of the Faster Region-Based CNN (RCNN) trained with the help of the YOLO V3 algorithm, find comparisons of the false positive and false negative rates when the real-time pictures are added to the ImageNet dataset. It is through this upgrade that the visual surveillance systems can now effectively detect guns. It was discovered that YOLO V3 was faster in comparison to Quicker RCNN in an actual real-time situation when applied on four different movies. In the case of weapons such as knives or guns that are captured by cameras, an algorithm developed [14] may update a security staff or operator automatically. The approach has a sensitivity of 81.18 percent and a specificity of 94.93 percent in the detection of knives so that it can bring false alarms to the minimal possible. The fire alarm system of the movie is highly sensitive (34.98 percent) and highly specific (96.69 percent). Histogram of Directed Tracklets [15] is a strong video classification method that can identify a problem in complex sequence and has a lot of potential to detect an anomaly. Tracklets are other options of tracklets than optical flow which only takes into account two consecutive frames edge qualities. New techniques have been created to enhance the descriptors of long-range motion projections. The possibility to statistically collect in two dimensions (space and time) cuboid video sequences is through their tracklets.

Third Partition: Arsenal detecting database.

To achieve a fair result in recognizing and identifying actual weapon, one has to capture the high-quality photos of the guns in many different angles and in various ways as there is no established dataset of what is considered a weapon to be. In order to make the developed neural network model more precise, we have removed all the redundant information in each weapon image. Because of this, we examined all the weapon pictures that were downloaded and altered the appearances of each one using several software programs depending on their contents. These programs consist of a diverse number of techniques in which one can edit images including rotating, cropping, padding, masking and clearing of background.

III. Methodology

A good quality dataset to be used to train machine learning models is paramount in each application. In this connection, we were forced to go through several Google images manually. The approximately eighty photos comprised an abundance of shots of all types of weapons. Creation of personal database with the help of Google images is brilliant. The photographs were then put in images folder. Further extensions will complicate the training, and the probability of mistakes will increase; the format of the image storage that is crucial is the jpg. On the contrary, with batch processing, all the photographs of the CNN are processed by being cropped to 416 x 416 pixels prior to training.

The main principles of computer vision and object detection are to find the objects content within digital images. The capacity to perceive objects has been enhanced significantly due to the occurrences of deep learning. YOLO may be considered as an object detector in its fundamental sense, trained. One of the most striking applications of convolutional neural networks (CNNs) is that they are able to learn bias and weights and generalize them to different objects and attributes in a given image. The CNN model is capable of extracting edges, and other high-level features of the input image by the application of the convolutional layer. These feature maps demonstrate the existence of input-detected features.

IV. RESULTS

The precise and detailed analysis of the findings of the three methods of locating firearms in video frames is displayed in Table I. The results in classification are through categorization. The above results indicate that our gun identification procedure on live surveillance footage is delivering the required outcomes.

A. CNN (Convolutional Neural Network)

C.N.R. is a regularly used acronym in image processing field. Changing the convolutional neural network input/output layer densities can enhance performance of a convolutional neural network. Attributes are identified by the use of layers. They are produced through applying a number of filtering algorithms to the input data to create the feature maps. Processing starts with a step where by the input values are multiplied with weights of each filter. The output is received by an activation function which could be the sigmoid or the Rectified Linear Unit (ReLU). A loss function is employed to assess the weights. Filtering takes place with feature maps in order to indicate some input items.

The decision of what CNN to use best identifies the best CNN to the job taking into account how the Faster R-CNN core architecture is going to be implemented. Many CNNs operate with a well-designed architecture to enhance accuracy but it is less cost-effective in terms of computing costs. Other designs can be fabricated into better models that can be used with embedded devices, but this approach still compromises accuracy.

Among the uses of pictures is to categorize one object in a picture. But in order to find something in an image, it is necessary to find it and make a growing box around it. illustrated a cartoon character with the power to recognize a gun. The shape of the detection kernel depends on the form of a 1 x 1 x (=bb x (4 + 1 + nc)) matrix. Based on this argument we would have; nc which corresponds to class count, bb which corresponds to bound box count, "4" which corresponds to bounded box coordinates and 1 which represents object confidence. Strides of 32, 16, and 8 are used in order to down-sample the input image to make three-scale predictions. The offered spot, the levels of certainty, and categorization errors are the three components in this case.

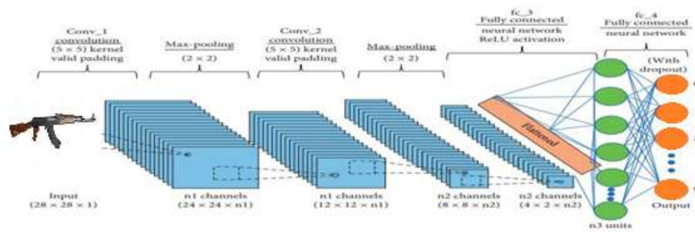

Figure 1. Deep neural network (DNN)

The deep neural network, as represented in Fig. 1, does the processing on the input image using various convolutional layers, activation, and pooling operations, which help to extract meaningful features. These attributes are then flattened and fed into fully connected layers to be classified. Lastly, the output layer generates the estimated value of the learned patterns.

B. YOLO Algorithm

YOLO being a Convolutional Neural Network (CNN), does not have difficulties with detection of objects. The Fig.2 below provides an example of how a convolutional neural network (CNN) can process incoming photos in the form of organised arrays of data and identify trends in it. YOLO does not compromise accuracy, being faster than its rivals. According to the model, it is able to view the big picture when performing an evaluation hence it considers everything when making predictions. Likely to other methodologies of convolutional neural networks, the YOLO methodology relies on the similarity between classes to rank regions.

When one area works very well, then this implies that it was selected to be described in the best manner in that category. The item identification within a real-life traffic stream can be one of the applications of YOLO. The processing rate is the primary advantage; YOLO is capable of the 45 frames per second rate. Check that out! YOLO depends on a convolutional neural network (CNN) to perform an overall analysis in one pass across an image and come up with a prediction, as opposed to the older approach of denoting a sliding window. Yolov3 utilizes a convolutional neural network (CNN) that was created by Alexey Darknet53 to perform image classification activities in order to achieve object recognition. The Transfer Learning method is required to

introduce new layers in an already existing model. It is surprising to nothing that the pre-trained weights of darknet53. conv.74 has been downloaded. Rather than relying on the weights that are randomly generated we will rely on these trained weights to train our model. This will result in an incredibly reduced number of time and computational resources required.
The problem of the failure of YOLO v2 to recognize small objects has been established. Complex information was lost as the layers reduced the resolution of the input. The second iteration uses an identity mapping with feature mappings of a prior layer; this is the case with the low-level feature acquisition. The issues, which are absent in the development of YOLO v2, include the lack of several features that are common in contemporary algorithms. The first models did not have any up sampling or skip connection as well as residual block. However, you have all of these features in YOLO v3. The fact that there is a possibility of identifying small details is indicated by the data points of Fig.2. The advances that were implemented to YOLO in version 2 and 3

TABLE I: YOLO V3, FASTER RCNN AND CNN VGG

SNO	MODELS	DATASET	ACCURACY (%)
1	Trained model YOLO V3	Video or Image dataset collected for present research	98.88
2	Faster RCNN (23)	Stream video	95.6
3	CNN VGG-16 (21)	IMDB	93

The comparison of the performance of the various models as indicated in Table I shows that the trained YOLO V3 model has the highest accuracy of 98.88% on the collected data. Faster R-CNN has a competitive performance of 95.6% accuracy on streaming video data. Comparatively, CNN VGG-16 model has an accuracy of 93% on the IMDB data set. This shows that YOLO V3 is more accurate when it comes to detecting weapons as compared to the other models.

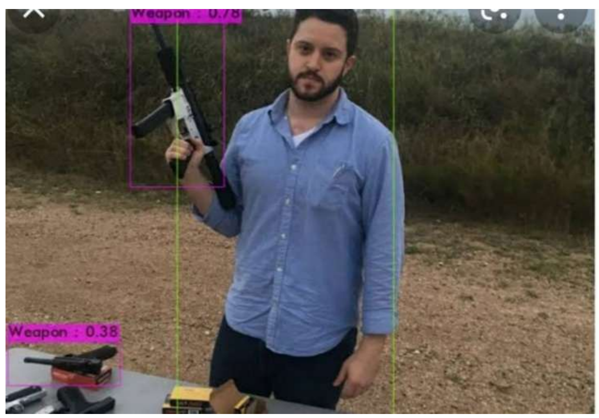

Figure. 2. Gun type of weapon identified objects

The system identifies and identifies the weapons in real-time by a detection model of objects as illustrated in Fig. 2. Bounding boxes and confidence scores are used to identify the person and objects like a rifle and a handgun. Weapons that are detected are also noticeable in the model making it easy to recognize and track. This shows how deep learning methods can be efficient in improving security and surveillance applications.

Figure 3. Category of weapon-detected objects: GUN

The system correctly identifies and classifies a weapon as a gun with a high score of confidence as shown in Fig. 3. The object is also indicated with a bounding box and marked as such and this indicates accurate localization and recognition. It means that the model can be successfully used to determine certain types of weapons during real-time situations.

Figure 4. Gun type of weapon identified objects

The system identifies a person with a gun and identifies the person and the weapon as indicated in Fig. 4 with the help of bounding boxes. The labels of person with gun and gun point to successful identification of object and its context. This illustrates that the model is able to identify threats involving weapons in complicated real-life situations.

Figure 5. The whole outcome of identifying weapons

Fig. 5 illustrates that the system can identify and recognize various weapons in an image by applying bounding boxes and classification labels. Several categories of firearms are well identified with the appropriate level of confidence, showing that the model can be used to deal with complex scenes containing several objects. This shows that the detection system is applicable in the large scale weapon identification tasks.

V. Conclusion and Future Scope

The capability to recognize armed people automatically in the security camera video has become of paramount importance in the high-crime areas nowadays. We must be in a position to identify and recognize firearms in case we want to halt criminal acts and ensure the correct individuals may respond. Most criminal activities involve the use or possession of weapons either whilst they are on the hand or when used. Possession or lack of a handgun is part and parcel of most criminal activities such as terrorism, illegal hunting and robbery.

Our trained CNN object recognition models and state-of-the-art YOLO V3 were used to detect weapons. With the assistance of our innovative method, machines will be able to identify dangerous weapons which might be through and inform a human operator when the need arises. We can tell that the YOLO V3 model is better compared to the YOLO V2 model to use a lesser number of computational resources, based on the experiments. Additional funding is needed immediately to prove that our human operators are performing well and they are increasing our monitoring capabilities. The current system will be eliminated with the advent of intelligent surveillance systems, which will be achieved with the advancement of low-cost storage, video infrastructure and the increased video processing capabilities. The use of robotic digital monitors will completely replace the old ways of surveillance when not only cheap computers will be available to the general population, but also rapid video processing and high-tech devices. In the case of surveillance systems, it is essential to be able to identify and report on things as quickly as possible. Better than its predecessors, this strategy could accomplish the same objective with minimal effort.

Be informed about the most recent news in the sphere! The future study should be in infrastructural aspects where continual research to incorporate autonomous robotic units in monitoring infrastructural areas, appraisal, and sending alerts to security personnel would be practiced with the aim of improving classification in security-related systems and allowing real-time analyses of data in such a way that people can interpret and utilize it. We shall know what guns are coated.

REFERENCES

[1] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," Advances in Neural Information Processing Systems (NIPS), 2012.

[2] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," International Conference on Learning Representations (ICLR), 2015.

[3] D. A. Forsyth and J. Ponce, Computer Vision: A Modern Approach, 2nd ed. Pearson, 2011.

[4] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, real-time object detection," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.

[5] R. Szeliski, Computer Vision: Algorithms and Applications. Springer, 2010.

[6] W. Stallings, Effective Cybersecurity: A Guide to Using Best Practices and Standards. Addison-Wesley, 2018.

[7] A. Hampapur, L. Brown, J. Connell, et al., "Smart video surveillance: Exploring the concept of multiscale spatiotemporal tracking," IEEE Signal Processing Magazine, vol. 22, no. 2, pp. 38–51, 2005.

[8] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. MIT Press, 2016.

[9] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2005.

[10] C. C. Aggarwal, Data Mining: The Textbook. Springer, 2015.

[11] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up? Sentiment classification using machine learning techniques," Proceedings of the ACL Conference, 2002.

[12] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," arXiv preprint arXiv:1804.02767, 2018.

[13] A. Warsi, M. Abdullah, and S. Khan, "Real-time weapon detection using deep learning for smart surveillance," International Journal of Advanced Computer Science and Applications, vol. 11, no. 4, pp. 345–352, 2020.

[14] M. Grega, A. Matiolański, P. Guzik, and M. Leszczuk, "Automated detection of firearms and knives in a CCTV image," Sensors, vol. 16, no. 1, pp. 47, 2016.

[15] H. Mousavi, A. Diba, and L. Van Gool, "Deep tracking: Seeing beyond seeing using recurrent neural networks," AAAI Conference on Artificial Intelligence, 2017.