A Comparative Analysis of Weapons Detection Using Various Deep Learning Techniques.

Sayma Tamboli, Komal Jagadale, Shreyas Mandavkar, Nitish Katkade, Taranpreet Singh Ruprah

Department of CSE, K.E. Society's Rajarambapu Institute of Technology, Urun, Islampur, Maharashtra, India.

sayma.tamboli786@gmail.com komaljagadale2017@gmail.com shreyash966977@gmail.com nitishkatkade94@gmail.com taranpreet.singh@ritindia.edu

Abstract— ***In crime scene analysis, object detection can be used to identify and track objects and people, which can help investigators to recreate the events and understand the sequence of actions that took place during a crime. The images are manually annotated, which is a process where a human expert goes through each image and marks the location and class of objects within the image. This process is important for training object detection algorithms as it provides the necessary ground truth data for the algorithm to learn from. In the case of crime scene analysis, high accuracy is crucial, as it can help ensure that no evidence is missed, but speed is also important, as time is often a critical factor in investigations. The proposed system in this study attempts to balance this trade-off by using algorithms like YOLOv5, SSD, and RCNN, which are known for their real-time performance while maintaining a high accuracy level.***

Keywords: Deep Learning, You Only Look Once v5(YOLOv5), Region-based Convolutional Neural Network(RCNN), Single Shot Detector(SSD), Object Detection

I. INTRODUCTION

Assault is a serious crime that can have severe consequences for the perpetrator. It is important to take steps to prevent an assault from occurring, particularly in public places where it can be more difficult to escape or get help. It is certainly important to notify the police as soon as possible when a crime.

Weapon detection is an important issue for public safety and security. By using deep learning algorithms, it is possible to develop models that can accurately identify weapons in real-time. These algorithms can analyze large amounts of data and identify patterns or features that may not be immediately apparent to humans. It involves identifying the presence of weapons, such as guns or knives, in a given area or location. This can be done using an image or video analysis, where the algorithm is trained on a dataset of images or videos that include weapons and other objects.

The proposed system aims to assist the police in detecting and identifying weapons in a variety of settings, including outdoor scenes. By using machine learning algorithms to analyze visual information, the system can potentially help the police to more efficiently and accurately identify weapons. This could be a valuable tool for law enforcement, as it could help them to more quickly and effectively respond to situations involving weapons, and potentially prevent crimes from occurring.

II. RELATED WORKS

The work by Laurie et al. [1] covered a real-time object recognition system that employed deep learning to identify objects in an image. This article mainly focused on the use of convolutional neural networks for object detection, which is the core technology used in YOLO architecture. The model was able to recognize items in photos using this technique with maximum accuracy and little latency. The use of depth-separable convolutional layers in this paper increased object detection accuracy while lowering computational model expenses. The feature extraction layer, the bounding box prediction layer, and the classification layer are just a few of the elements that make up the YOLO architecture as well as the technique that it employs. However, if the training data is biased or limited, the model's performance may not be as accurate on new, unseen data.

The work by Redmon et al. [2] discussed the architecture of YOLO, a real-time object detection system capable of recognizing objects in an image with high accuracy and low latency. This study focused on various components of the YOLO architecture such as a feature extraction layer, a bounding box prediction layer, and a classification layer. The article also discussed the use of convolutional neural networks for object detection. The aim of the project was to investigate the use of Transfer Learning using Convolutional Neural Networks (CNN) to improve the accuracy of an object detection model and its application to new image datasets. The study did not discuss any potential ethical considerations related to the use of object detection models, such as privacy concerns or potential biases in the training data. Therefore, the ethical implications of the study were not fully explored.

The research article by Fan et al. [3] provided an enhanced pedestrian detection system based on the SSD model of object detection. As an extra layer to the SSD concept, this system featured the Squeeze-and-Excitation (SE) model. The SE model allowed the system to learn from its flaws and enhance the precision of small-scale pedestrian identification. The enhanced model's accuracy was tested using the INRIA dataset. This study served as a resource for comprehending the SSD concept and future applications.

Several works have shown that CNNs can also be used as feature extractors for various vision tasks.
Razavian et al. [4] proposed an approach to extract CNN features from an off-the-shelf network to perform object recognition. Similarly, Donahue et al. [5] used the activation of convolutional layers to improve object recognition accuracy. In another work, Oquab et al. [6] used pre-trained CNNs for image retrieval tasks.
In order to anticipate the hourly wind speed, Ling Chen and Xu Lai (2011) [7] compared the experimental results generated using an Artificial Neural Network (ANN) and Autoregressive Integrated Moving Average (ARIMA). Comparatively speaking, the ANN model outperforms the ARIMA model. However, in order to achieve better outcomes, integrated models or a numerical weather forecasting method are required when the variation is too big.

Shiju Sathyadevan, Devan M. S., et al. (2014) [8] identified places with a high possibility of experiencing crime and depicted crime-prone zones. The authors used the Naive Bayes classifiers algorithm, a supervised learning and statistical approach for classification that has a 90% accuracy rate, to classify the data. However, the study covers crime data only from the years 2000–2014 and therefore, the findings may not be applicable to more recent time periods.
Dr. N. Geetha, Akash Kumar. K. S, Akshita. B. P, Arjun. M (2021) [9] demonstrates a technique for automatically spotting firearms in the video, which is useful for monitoring and exercising control. The You Look Only Once (YOLOv3) algorithm was implemented to find weapons in real-time footage. In conclusion, the yolov3 algorithm outperforms the earlier CNN, R-CNN, and faster CNN algorithms in terms of speed.

Neil Shah, Nandish Bhagat, and Manan Shah (2021) [10] aimed to study how law enforcement agencies or authorities might employ a mix of ML and computer vision to detect, prevent, and solve crimes much more accurately and quickly. In conclusion, they have mentioned that Machine Learning and computer vision approaches can help law enforcement agencies advance.

III.METHODOLOGY

A) Dataset:
The dataset is used to train the model. Raw images often need to be pre-processed before they can be used for further analysis or model training. The YOLO v5, SSD, and R-CNN models were trained on a publicly available weapon image dataset provided by the University of Granada research group, with a specific split of 2208 training images, 660 testing images, and 174 validation images. The aim of the training was to detect pistols, rifles, and knives, and the number of training epochs was set to 100. This dataset is annotated using roboflow which is an online tool to handle annotation formats. Rectangular bounding boxes were used to annotate the images manually. The dataset was divided into 3 classes (i.e., weapon, knife, and person) and each class included an almost equal number of images.

B) Separation of the dataset into train and test data:
The dataset was divided into 3 classes (i.e., weapon, knife, and person) and each class included an almost equal number of images. It will generate a text file for each image which consists of coordinates used for markings. The label file will store the classes for which the images will be marked. You can then upload the zip file to a cloud storage service such as Google Drive and then you can split it utilizing the chosen split ratio, into two sets: a training set and a test set.

C) Workflow Diagram:

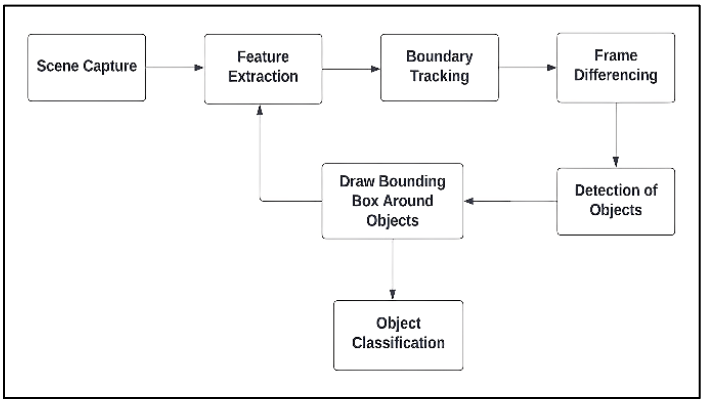

Fig.1 Explains the workflow of the system.

1. Scene Capture: This is the initial stage where it must capture a view of a real environment that is most probably a crime scene

2. Feature Extraction: Extraction of features is done from the initial dataset; these new features have this little redundant information in them and therefore features are often less intuitive to understand. Feature extraction is achieved by using a series of convolutional layers to identify features at different scales and resolutions.

3. Boundary Tracking: In object detection, boundary tracking refers to the process of identifying and locating objects in video sequences or images.

4. Frame Differencing: Frame differencing is a method for finding objects to detect changes in a video stream by subtracting the current frame from a previous frame.

5. Detection of Object: The model uses a combination of convolutional neural networks, region proposal algorithms, and classification and regression models

to accurately detect and classify objects within the input data.

6. Draw a Bounding Box Around the Object: Bounding Boxes act as the reference points for weapon detection and they are normally in rectangular shape. The rectangles should be drawn over images and then must identify X and Y coordinates for each image.

7. Object Classification: The dataset contains three different classes i.e., knife, person, and guns so will determine which object from these three classes is present in the given scene or image it refers to the model to find out which class is present.

D) Pre-processing and Augmentations:
Data augmentation can be a useful technique to enhance the performance of a deep learning model by increasing the size and diversity of the dataset, which can lead to better learning and generalization. Therefore, the size and diversity of the dataset can be increased by introducing variations into the data by implementing data augmentations.

From the information provided, it seems that the project explored several data augmentation techniques to improve the performance of the model. The techniques like noise addition, color shifting, rotation, flipping, cropping, etc. were commonly used for data augmentation methods in deep learning to increase the diversity of the training dataset and prevent overfitting. By applying these techniques to the training data, the model is exposed to more variations of the same object, which helps to make it more robust and better at generalizing to new, unseen data. Auto-Orient helps to orient the image correctly, Resize adjusts the size of the image, Grayscale converts the image to grayscale, Auto-Adjust Contrast improves the contrast of the image, and Outputs per training example refers to the number of bounding box predictions made per image. Rotation and Shear add rotational and shearing transformations to the image, respectively, which help to increase the diversity of the data even further.

IV.IMPLEMENTATION

A) Algorithms

1. In YOLOv5, these were the two unified blocks that turned into a single monolithic block.
1. feature extraction
2. object localization
YOLOv5 has three main components namely Backbone, Neck, and Head.

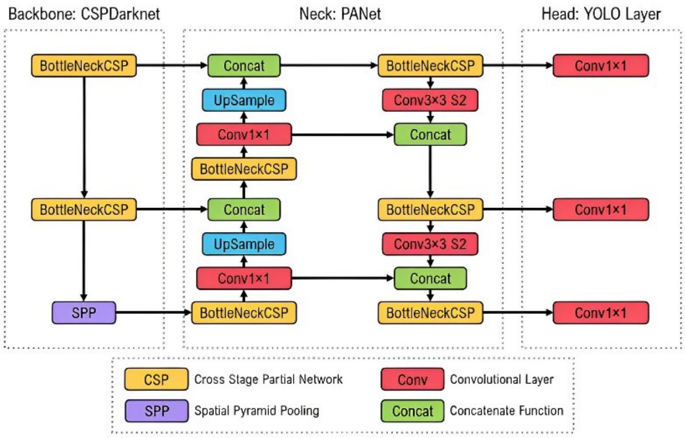

Fig.2. YOLOv5 Architecture [9]

YOLOv5 employs Cross Stage Partial Networks for the purpose of obtaining instructive information from the input image (CSP). The model neck creates feature pyramids (FP). Anchor boxes are used to apply feature class probabilities. In YOLOv5, hyperparameters are used to control the model's architecture, training process, and performance. For YOLOv5 the hyperparameters were 50 along with the Batch size and Learning rate as 16 and 0.001.

2. The object detection technique known as Region-based Convolutional Neural Network (RCNN) is based on the visual data of images.

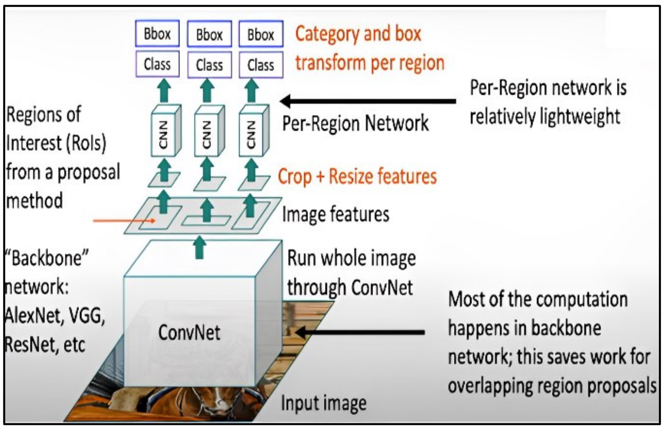

Fig.3 R-CNN Architecture [10]

Figure 3, shown above, describes the RCNN architecture. After determining the region where objects might be present, then extract the feature of the input image provided using a feature extractor. The hyperparameters for the R-CNN algorithm were 100 with a similar batch size and learning rate as the YOLOv5 algorithm.

The following steps are often used by R-CNN to classify objects.
1. Take note of the n areas (Region Proposals) where the original image's items will be located.
2. Use CNN to extract choices from regions that are 227x227 (AlexNet) or 224x224 (VGG16) in size.

3. The 4096-dimensional Ultimate Output Layer (UOL) forecasts class adjustments using SVMs and prediction boxes.

3. The SSD architecture is depicted in the picture and comprises additional layers that are built on top of a base CNN network, such as VGG or MobileNet.

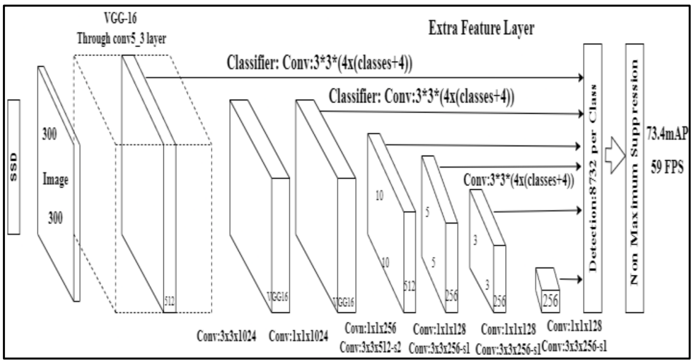

Fig.4. SSD VGG-16 Architecture [11]

The SSD technique extracts information from each grid cell using a sequence of convolutional and pooling layers. Each grid cell is then subjected to a classifier to forecast. The hyperparameters for the SSD algorithm were 150 with a similar batch size and learning rate as the YOLOv5 algorithm.

B) Performance Metrics

I. True Positive Results

This suggests that the model is performing well in recognizing small and blurry objects. The high accuracy on these objects indicates that the model has effectively learned the relevant features and patterns needed to correctly identify them.

Fig.5. Image for True Positive Result

II. False Positive Results

A false positive is a prediction made by the model that an object is present in an image when in reality it is not this can result in incorrect or misleading results and is a common challenge in object detection tasks a high number of false positives can negatively impact the Precision of the model and reduced its overall performance.

Fig.6. Image for False Positive Result

III. True Negative Results

True negative is a Precision made by a model that there is no object in an image when in reality there is not a high number of two negative indicates that the model is able to accurately identify when there is no object present in the image which is important for awarding false alarms and improving the precision of the model.

Fig.7. Image for True Negative Result

C) Evaluation Metrics

AP (Average Precision) is a commonly used metric in object detection to evaluate the performance of the model. It considers both precision and recall and calculates the average precision for a range of recall values. The higher the AP score, the better the model is at correctly identifying objects in an image.

V.RESULT

Comparative Analysis Study of algorithms:

Here are the derived results for the comparative analysis study of YOLOv5, RCNN, and SSD algorithms.

1) You Only Look Once (YOLO V5):

The metrics curves during the training process are displayed in Figure 8. Upon evaluation, the YOLO5 model attained precision and recall scores for validation, along with mAP scores at @0.5IOU and @0.95IOU. These results validate the efficacy of the approach in accurately predicting signs in various environments.

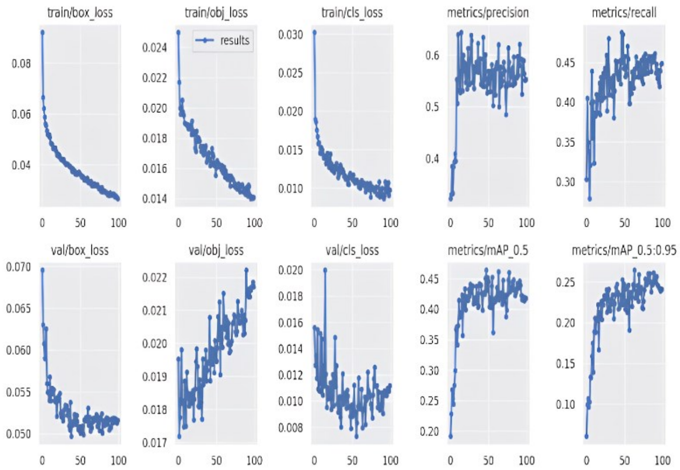

Fig.8. You Only Look Once model evaluation

A confusion matrix consists of a table that compares the predicted and actual labels for a set of data points. The rows represent the predicted labels, while the columns represent the actual labels. The entries in the table count the number of data points that fall into each possible combination of predicted and actual labels.

By analyzing the confusion matrix, it can gain insight into the model's performance, including which classes are being classified correctly and which are misclassified. This information can then be used to adjust the model's parameters, such as its architecture or hyperparameters, to improve its performance. The confusion matrix is displayed in figures 9,11 & 13.

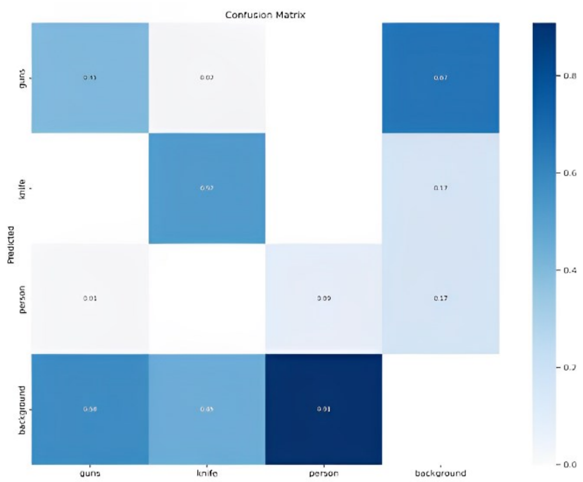

Fig.9. Confusion matrix for YOLO V5

2) Region-based Convolutional Neural Network (RCNN):

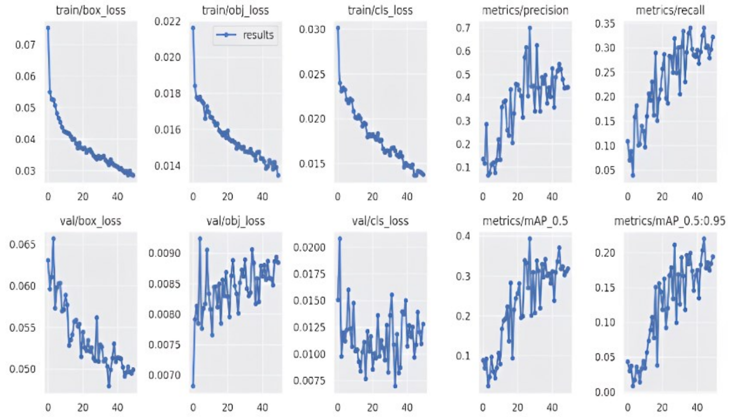

Fig.10. Region-based Convolutional Neural Network model evaluation

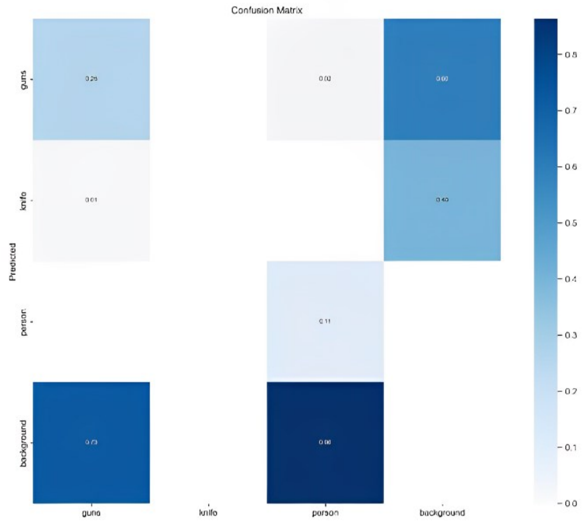

Fig.11 Confusion matrix for Region-based Convolutional Neural Network

3) Single Shot Detector (SSD):

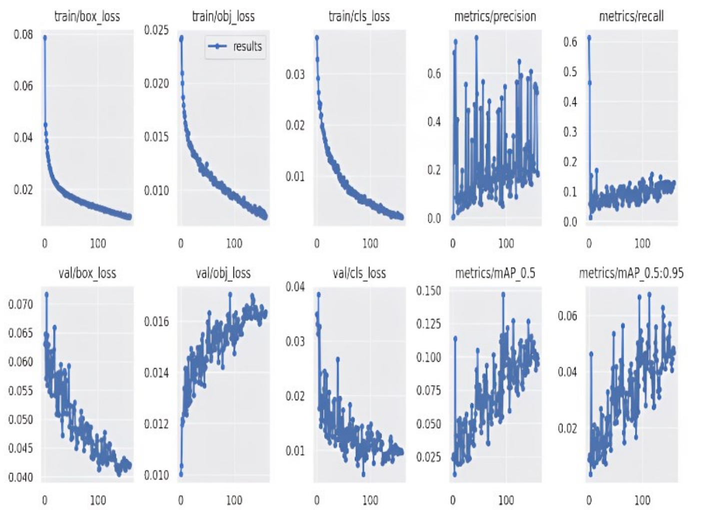

Fig.12. Single Shot Detector model evaluation

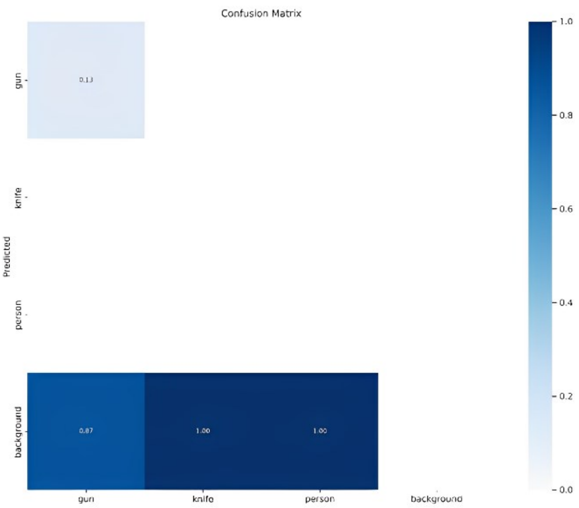

Fig.13.Confusion matrix for Single Shot Detector (SSD)

The YOLOv5 model was used to perform object detection on video footage and the results were evaluated in terms of true positive detections.

Table 1. Comparative analysis of map values

MODELS	YOLOv5	RCNN	SSD
Map value (Mean Average Precision Value)	56.2	47.1	36.7

From the information provided in table 1. Comparative analysis of map values, it seems that the model performed well in detecting small and blurry objects, as evidenced by a high mAP (mean average precision) score of 56% for small objects. This suggests that the YOLOv5 is known for its robustness in accurately detecting small and difficult-to-see objects in video footage. Similarly, the SSD model is known for its ability to perform well in detecting small and blurry objects in video footage, as indicated by the high mAP score of 37% for small objects. The RCNN model performed well in detecting small and blurry objects in video footage, as indicated by the high mAP score of 47% for small objects.

VI.CONCLUSION

Different models as YOLOv5, RCNN, and SSD were used for image processing and computer vision, and for comparing their performance on the dataset. Thus, a comparative analysis of these algorithms was necessary for the development and improvement of weapon detection systems, as it allows for the identification of the most effective and practical methods for detecting weapons in a variety of different scenarios. As mentioned YOLOv5 has the best prediction accuracy, but it was slower than the other two models.YOLOv5 model included an affine-tuning approach to optimize its performance, and YOLOv5 has improved compared to previous versions and RCNN can be more accurate but slower. Ultimately, the choice between YOLOv5 and RCNN depends on the specific use case and the trade-off between accuracy and speed that is acceptable for the application.

REFERENCES

[1] Reagan L. Galvez; Argel A. Bandala; Elmer P. Dadios; Ryan Rhay P. Vicerra; Jose Martin Z. Maningo: TENCON 2018 - 2019 IEEE Region 10 Conference.

[2] Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi: You Only Look Once: Unified, Real-Time Object Detection: IEEE conference on computer vision and pattern recognition (2016).

[3] Wei Fan, Zhenying Xu, Ziqian Wu: Improved SSD-assisted algorithm for surface defect detection of electromagnetic luminescence(2021).

[4] Razavian, A.S., Sullivan, J., Carlsson, S., Maki, A.: Visual instance retrieval with deep convolutional networks. arXiv preprint arXiv:1412.6574 (2014).

[5] Jeff Donahue, Yangqing Jia, Oriol Vinyals, Judy Hoffman, Ning Zhang, Eric Tzeng, Trevor Darrell: DeCAF: A Deep Convolutional Activation Feature for Generic Visual Recognition: arXiv preprint arXiv:1310.1531 (2014).

[6] Oquab, Maxime, Bottou, Léon, Laptev, Ivan, Sivic, Josef, et al. Weakly supervised object recognition with convolutional neural networks. In NIPS, 2014.

[7] Chen, Ling, and Xu Lai. "Comparison between ARIMA and ANN models used in short-term wind speed forecasting." Power and Energy Engineering Conference (APPEEC), 2011 Asia Pacific. IEEE, 2011.

[8] Sivaranjani, S., S. Sivakumari, and M. Aasha. "Crime prediction and forecasting in Tamil Nadu using clustering approaches." Emerging Technological Trends (ICETT), International Conference on. IEEE, 2016.

[9] Dr. N. Geetha, Akash Kumar. K. S, Akshita. B. P, Arjun. M: Weapon Detection in Surveillance System: International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181 IJERTV10IS050077 Vol. 10 Issue 05, May-2021.

[10] Neil Shah, Nandish Bhagat and Manan Shah: Crime forecasting: a machine learning and computer vision approach to crime prediction and prevention(2021)

[11] MalavikaNair, Mathew Gillroy, Neethu Jose, and Jasmy Davies in their “i-surveillance Crime Monitoring and Prevention Using Neural Networks”, International Research

Journal of Engineering and Technology (IRJET), e-ISSN:2395-0056Volume: 05 03rd March 2018.

[12] jha P,jha R,Sharma A (2019) Behavior analysis and crime prediction using big data and machine learning.Int j Recent Technol Eng 8(1):461-468

[13] Harsh Jain, Aditya Vikram, Mohana, Ankit Kashyap, Ayush Jain Telecommunication Engineering, “Weapon detection using AI and deep learning for security application”, IEEE (2019).

[14] M. K. Mohamed, A. Taha and H. H. Zayed, “Automatic gun detection approach for video surveillance,” International Journal of Socio Technology and Knowledge Development, vol. 12, no. 1, pp. 49–66, 2020.

[15] J. Lim, M. Istiaque, A. Jobayer, V. M. Baskaran, J. M. Lim et al., “Gun detection in surveillance videos using deep neural networks,” in Proc. of APSIPA, Lanzhou, China, vol. 1, no. 1, pp. 1–6, 2019.

[16] Erssa Arif, Syed Khuram Shahzad, Muhammad Waseem Iqbal, Muhammad Arfan Jaffar, Abdullah S. Alshahrani and Ahmed Alghamdi: Automatic Detection of Weapons in Surveillance Cameras Using Efficient-Net (2022).

[17] Karabo Jenga, Cagatay Catal, Gorkem Kar: Machine learning in crime prediction: Journal of Ambient Intelligence and Humanized Computing (2023).