2021 2nd International Conference for Emerging Technology (INCET)
Belgaum, India. May 21-23, 2021

Multiclass Weapon Detection using Multi Contrast Convolutional Neural Networks and Faster Region-Based Convolutional Neural Networks

Rahul Reddy
School of Electronics and Communication Engineering
Vellore Institute of Technology,
Vellore, India
vaderahul9@gmail.com

Gyan Vallabh K
School of Electronics and Communication Engineering
Vellore Institute of Technology,
Vellore, India
gyani1910@gmail.com

Sai Sharan
School of Electronics and Communication Engineering
Vellore Institute of Technology,
Vellore, India
saisharan68@gmail.com

2021 2nd International Conference for Emerging Technology (INCET) | 978-1-7281-7029-9/20/$31.00 ©2021 IEEE | DOI: 10.1109/INCET51464.2021.9456407

***Abstract*— In this day and age, increasingly easy access to firearms and other hand-held weapons has stirred up public violence concerns. Many of these weapons are comfortably concealed. The tremendous developments in technology can assure more security in public places by detecting such weapons in real-time and alerting the concerned authorities before any damage. In this paper, we propose using two state-of-the-art algorithms for performing real-time detection of concealed hand-held weapons. The algorithms used are Multi Contrast Convolutional Neural Networks (MC-CNN) and Faster Region-Based Convolutional Neural Networks (Faster R-CNN). Additionally, we present a comprehensive comparative analysis and evaluation of the Faster R-CNN and MC-CNN in detecting the weapons. This study has diverse industrial applications in real-time bank surveillance cameras and other public places, provided that they are under CCTV surveillance.**

Keywords - MC-CNN, Faster R-CNN, CNN, Multi-Class Weapon Detection, Deep Learning

I. Introduction

Nowadays, it has become easy to access weapons, especially for people in countries where weapons ownership is legal. The homicide incidents have increased rapidly in recent years. This raises extreme concern about public safety, which should be taken into utmost consideration. The number of deaths annually due to gun violence in the US itself is around 1220. Though security personnel are being appointed at public places, it is tedious to deliberately and manually monitor everyone in a crowd.

Moreover, in many cases, even if the security responds, it would be too late by then. There is a need to alert the authorities immediately if any standard weapons are frequently used in criminal activities. To avoid casualties, detection must be performed accurately and quickly in real-time. One of the promising solutions for this problem is to deploy surveillance cameras with inbuilt weapon detection and alerting algorithms. Since CCTV surveillance is becoming more common in most places such as banks, Jewellery stores, Railway stations, etc., implementation of detection techniques through these cameras will be cost-effective and fruitful.

In recent years deep learning has established a landmark in object detection, classification, and image segmentation by producing productive results. Although existing studies currently perform weapon detection, they are hindered by various limitations such as ineffective real-time processing, the noise of weapon images, etc. Therefore, in this paper, we introduce two ways to detect weapons in public places through CCTV cameras automatically. In this research work, we carefully scrutinize MC-CNN and the Faster R-CNN models used for weapon detection. Additionally, we present a detailed comparative study delineating the use case of each model.

Weapon detection is one of the interesting applications of computer vision. Our paper involves image recognition and segmentation. The recognition outputs a classification label, and image segmentation creates a pixel-level understanding of objects in the scene, which allows us to track the weapons in the frame. a Faster R-CNN model is widely used for object detection and tracking. It is composed of a feature extraction network which is typically a pre-trained CNN. Then we will be training two subnetworks of Faster R-CNN on our contrast-enhanced and augmented dataset. The first is the Region Proposal Network (RPN), which generates object proposals, and the latter used to predict the precise class of objects. In the end, we use ROI pooling, an upstream classifier, and a bounding box regressor which displays whether a weapon is detected or not.

An MC-CNN (Multi-contrast CNN) model consists of individual CNNs trained on differently pre-processed data. Our training phase involves contrast enhancement of the greyscale images to improve accuracy, followed by the data augmentation process so that enough data is available to train the model efficiently without overfitting. In the end average of all the individual CNN predictions is taken as an output.

The structure of the paper is as follows. Section 1 contains the introduction. Then we move on to Section 2, which is the related works followed by an overview in Section 3. Section 4 includes the proposed models. Results and description can be found in Section 5, and we end with a conclusion.

II. Related Work

Today, hand-held weapons such as pistols, revolvers, etc., are primarily used for criminal activities. Moreover, easy access to these weapons has been raising concern among people in public areas. The surveillance cameras and security personnel have not been effective in diminishing the high risk of criminal activities. A possible solution is by incorporating deep learning algorithms with CCTV cameras to detect hand-held weapons in real-time automatically. This could effectively reduce the crime rate by intimating the concerned authority/security before the loss, whenever a weapon is detected. Although there are pre-existing object detection techniques, most of them are based on images that need low-level highlights and artifact features. This doesn't allow for an optimal solution. So there is a need for reliable and robust techniques that can reduce the processing time and produce accurate results. In this research work, we will be thoroughly analysing RCCN and MC-CNN for detecting hand-held weapons.

Weapon detection can be implemented using various approaches; however, only the most robust techniques provide promising results. In [1] by Rohit Kumar et al., the authors propose a hybrid approach for weapon detection through color-based segmentation and interest point detector. A median filter is used for removing the noise from the video frames. Unrelated objects and colors were eliminated with the help of k mean clustering filled by the blob extraction. Then small gaps in the blob are eliminated by performing morphological closing, which is achieved by erosion of image with structuring element and then dilation with structuring element. Interest point identification by Harris corner detector and building descriptor of each interest point using FREAK feature extractor were two primary steps involved in feature extraction. This approach worked well in most situations. However, it produced poor performance in illumination changes because of inaccurate color-based segmentation.

Although multiple object detectors are used for detection, they are inefficient when capturing weapons at different scales and non-canonical datasets. To address these, in [2], Jun Yi Lim et al. proposed an approach in which a dataset is constructed with a focus on surveillance-based handguns and non-canonical datasets. They found out that the constructed dataset is relatively well suited for representation learning because of various factors such as large non-canonical handguns in realistic settings. The created dataset is trained by several models such as M2Det, localization losses, unconventional classification losses, etc. It was observed that Generalized Intersection over Union (GIoU) improved the model's performance by correcting base truth. Inconsistencies of predicted handguns due to low classification scores are addressed by implementing Adaptive Surveillance Image Partitioning (ASIP) in the model. It concluded that ASIP performed well in autonomously detecting handguns.

According to Dhiraj et al. in [3], automatic X-Ray screening and baggage testing for any weapons at the airport are still a significant concern because of various factors like the imbalance between false alarms missing detections, adaptiveness, and robustness. Here Dheraj et al. compared different deep learning techniques to perform object detection. The feasibility of the YOLOv2 model is also explored for detecting threat objects in the baggage screening process. Two approaches Faster-R-CNN and YOLOv2 were finalized considering their suitability and usage for multi-threat detection in baggage images. Efficient data augmentation is done by various operations of the data such as rotating images, flipping, skewing, random, distortion, and zooming. They compared the Fast R-CNN and YOLOv2 models and observed that Fast R-CNN produced better Accuracy, Recall, and precision performance than YOLOv2. In [4], Justin Lai et al. used Over feat network implementation to classify detection detect in images. For Baseline, the VGG-16 classifier pre-trained on ImageNet is used. Overfeat network is implemented through Tensorbox and trained by adjusting hyperparameters in different ways with different confidence thresholds and learning rates. By decreasing the confidence threshold, there was a rapid change of bounding box sizes while training. The author concluded that the drastic difference was directly related to the confidence threshold used.

Although CNN helps detect the weapons, there is always a scope to look out for more complex algorithms such as R-CNN, MC-CNN, YOLO, etc. In [5], Gyanendra K Verma et al. paper put forth significant gun challenges detection such as occlusion of a gun, the inter-class variation which occurs due to variation in color and structure, noises in the gun image, deformation of the gun, and need for real-time processing and presented an approach to detect the weapons from the cluttered scene using Convolutional Neural networks (CNN). CNN is implemented using MatConvNet without Graphical Processing Unit(GPU) for computer vision applications. ImageNet pre-trained model is used to train the system to reduce the training time. The system performance is tested for occlusion, the different backgrounds of guns, etc. The performance is evaluated in terms of false-positive rate (FPR), True positive rate (TPR), False Detection Rate (FDR), Positive prediction Rate (PPV). It is observed that the CNN method performed well with the inter-class variation and intraclass similarities.

Every possible detection might be confined to single object detection, but the difficulty always seems to arise when detecting multiple objects in real-time. Their dimension makes it even more challenging and virtually impossible to the detection of various objects. Weapons that mostly fall under this radar support the increase of problems at banks, ATMs, etc. In [6], Siham Tabik et al. proposed Object Detection Binary Classifiers methodology based on deep learning to identify small objects handled similarly: Application in video surveillance, the basic idea was to introduce to a new face of application with OVA (one verse all) and OVO (one versus one) which are involved in images detections supports existing algorithms in a better standard. In general, CNN picks up a faster rate involving these procedures. Despite having a very close approach to detecting these small objects, we still face misconceptions regarding the object's actual terms. Thus, introducing new concepts like VOTE random, VOTE by weight could bring us a definite solution to these problems. In the Sohasweapon database and ODeBiC methodology based on Deep learning, we have three finite steps which involve: detections, processing, comparison, and storing in database. Thus with all these consideration factors, we can understand how these tiny objects can be recognized.

As stated in [7], the idea of object detection is the same, whether it's static or dynamic. The only question is that how fast we can process and determine the objects. In Real-time gun detection in CCT, these setups are mainly placed in ATMs and banks to detect unusual activities. Now, we have two methods in consideration: CNN and R-CNN. Since detection of these guns should be done faster, we always have to adopt the process that yields faster results. Hence, R-

CNN was adopted by combining it with Feature Pyramid Network (FPN) and enhanced dataset with synthetic data. This improved the detection of small objects in real-time. These methods are still improving.

In [8], Harish Jain et al. presented an approach that addresses the issues of Weapon/ anomaly detection by using machine learning models such as Regional Convolutional Neural Network (R-CNN), Single Shot Detection (SSD). SSD algorithm achieved new records in terms of performance detection and precision. SSD brings forth default boxes and multiscale features to overcome the drop in accuracy, allowing SSD to match the Faster R-CNN's accuracy. These advancements helped in using lower resolution images which further pushes speed higher. Few setbacks of faster R-CNN compared to the SSD model are vast and unwieldy training data. In SSD, training occurred in more phases which made the algorithm faster. With faster R-CNN, it cannot provide accurate real-time detection due to time spent on region proposals. It is observed from the results that the pre-labelled dataset provided better accuracy in SSD and R-CNN since it is trained for a large number of images in comparison to a self-built dataset. It is concluded that although R-CNN and SSD performed well in weapon detection, there is a trade-off between speed and accuracy. SSD algorithm produced a relatively better rate, whereas Faster R-CNN instead produced better accuracy.

In [9], Roberto Olmos et al. proposes the Automation Handgun Detection Alarm in Videos using Deep Learning. Since we are accessing the videos, we require a good amount of storage or a new database varying in time. Deep CNN algorithm is an essential aspect of the process, alongside we have two effective methods: sliding window approach and Region proposal approach. Here the video is spilled (or considered as windows), and they are scaled to the maximum possible range at various locations so that the desired output is obtained. Simultaneously, the following method selects the exact regions of the video and applies the preferred detection method. This makes the detections of guns very effective.

Today, as we see, many object detections are basically through either CNN, R-CNN, or involve other Deep Learning techniques. Gradually image processing also made its efforts in the competition. According to Bhavana et al. [10], detection happens with the object stored in the database. Still, image processing converts the image into a discrete form. It narrows down the vision to the maximum possible extent and uses few mathematical tools to match the coordinates and compare with the original object. Initial generations use the energy that is emitted from the body in the detection of guns. Next generations used the frequency sent by the sensors to detect the objects. Finally, in the third generation, they combine all the senor aspects for processing object detection. In [11], most related work in any detections happens through CNN or the R-CNN algorithm. Still, Automated Detection of Firearms and Knives in CCTV images focuses mainly on using CCTV reordered data and tracking the object. The recordings are broken down into images, and these images are sorted through a sliding window algorithm. The extractions done from these images are applied to SVM classification, and hence the decisions are made whether the object needs to be flagged and if the authority has to be alerted or not. The problem during this process arises only in situations where the knives are misled with the offender or defender because the images taken from the reordering are based on fixed points by pausing the video.

MC-CNN is one of the most organized yet not widely well-known techniques. Its performance in detection is tremendous. In [12], Prajval Mohan et al. proposed Multi-contrast Convolutional Neural Networks (MC-CNN) and Convolutional Architecture for Fast Feature Embedding to detect defects in a tyre. The paper built a custom dataset with 528 images followed by a data augmentation process to increase the sample size. Once Image augmentation and Contrast normalization are done on the dataset, both the algorithms are tested on the dataset in terms of various parameters such as specificity, sensitivity, batch size, accuracy, etc. it can be inferred that the accuracy of MC-CNN increase with the number of nets of CNN used and use of automatic deep learning. The results indicate that the MC - CNN algorithm performed relatively well compared to Fast feature embedding in detecting the defects. The paper concludes that higher accuracy is directly attributed to the dataset's size and adaptive techniques used.

A. *Limitations*

It is clear from the previous works that R-CNN is based on a selective search approach to extract Regions Of Interest (ROI). Here ROI resembles a rectangle boundary of an object image. The number of ROIs varies depending on the scenario. Subsequently, each ROI is fed through Neural Networks to produce output features. For each of these features, a set of Support Vector Machine (SVM) classifiers is used to detect the object within the ROI. Although this works well for object detection, it falls short in some aspects, such as; the vast amount of time required to train the network, its static nature, which hinders the learning process. The two-stage pipeline(data needs to be trained by CNN followed by multiple SVMs) could be its disadvantage. Its expensive training requires not only more disk space but also reduces the computational speed.

For MC-CNN, its good performance is attributed to its delicate structure, which includes several CNN columns. However, MC-CNN has high computational and hardware costs compared to single CNN because of its complex system. Our paper presents a comprehensive comparative analysis of Faster R-CNN and MC-CNN on our custom-built data set.

III. Overview

This paper performs weapon detection on our custom-built dataset images by applying both Multi Contrast - CNN and Faster R-CNN. In the first case, we use Multi Contrast CNN. The method aims to train several CNN to detect the weapons and take the mean of the results to get the final result classification. In the latter case, the RPN generates a region of proposals from all the proposed images, and a fixed-length vector is extracted for each region using the ROI pooling layer. Finally, the extracted values are classified using Faster R-CNN.

A. *Motivation*

The current works use algorithms that work fine for detection in a static environment but can be impractical for real-time weapon detection. Most of the existing models on weapon detection have high training time and a tedious training process on the dataset. Additionally, the current works performed only binary classification, which produced “YES” if a weapon is detected, else “NO.” In this paper, we will be working on multi-class classification wherein we detect different categories of guns. Our work’s application scope is limited to guns and can also be used to detect other similar hand-held weapons, provided we have an appropriate dataset.

B. Dataset

Since there are no available multi-class datasets that are suitable for our model, so for this paper, we have created our own dataset. We have collected 788 images from google and divided them into four different gun classes, namely: Pistol (186), Rifle (208), Shotgun (193), and Sniper (201). The pictorial representation of each class can be seen in Table I. We have split the dataset in a ratio of 2:3 as the training and testing set, respectively. Our dataset will be publicly released after publication.

TABLE I. TABLE TYPE STYLES

Type of Weapon	Table Column Head
Riffle	
Pistol	
Shotgun	
Sniper	

C. Pre-processing

The dataset created by us includes images of different sizes and aspect ratios. Some of these are in greyscale. Therefore, we first perform data pre-processing on the dataset so that the images can be accurately used by Faster R-CNN and MC-CNN algorithms.

We have used the sliding window method to resize the images to 128 x 128 pixels each. This makes sure that the images are of similar size and quality as that of CCTV and surveillance cameras [11]. Since our dataset also contains greyscale images, we have performed contrast enhancement by making the best use of available colors on the output to make the images more apparent and more distinguishable. We have also used Image Augmentation to expand our dataset’s size since CNN could not perform well on smaller datasets.

Contrast Enhancement

It is an image processing technique to manipulate and redistribute the image pixels in the linear or nonlinear form to distinguish the images. (Eq. 1)

$$G(f) = \frac{\mathrm{f} - \mathrm{fk}}{\mathrm{fk} + 1 - \mathrm{fk}} * (\mathrm{gk} + 1 - \mathrm{gk}) + \mathrm{gk} \qquad (1)$$

For $f_k < f < f_{k+1}$, k=0, 1, …, k-1;
Contrast enhancement uses piece-wise linear stretching. In the above equation, g(t) is the output function. f(t) is the input function to which enhancements are altered.
K-parameters:
f_k is the starting position of the input {k=0, 1…, k-1};
g_k is the starting position of the output {k=0, 1…, k-1};

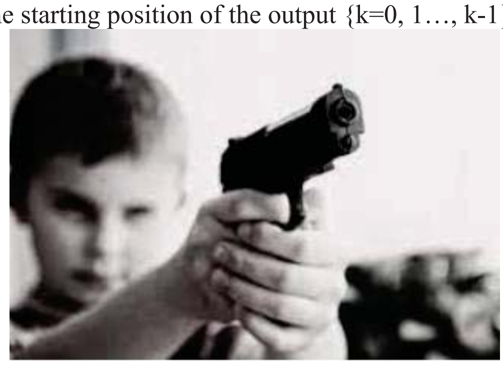

Fig. 1. Original Image

Fig. 2. Contrast-enhanced Image

It can be inferred from the above figures that Figure 1 is initially in greyscale, and after undergoing contrast enhancement in Figure 2, there is a significant improvement. Additionally, in Figure 2, we observe a sharper image.

Data Augmentation

Image augmentation consists of various image manipulations such as flips, shifts, zooms, etc. Its primary purpose is to expand the training set by adding new viable images by performing operations on an existing image.

D. Testing parameters

For the experimental analysis, we worked on an Intel i7 (8th generation) processor with a max clock speed of 2.6 GHz and a turbo speed up to 4.5Ghz. A 4 core NVIDIA GeForce GTX 1650 Max-Q dedicated graphic card was also included.

IV. PROPOSED MODELS

A. *Multi-Contrast Convolution Network*

A Multi-Contrast Convolutional Network comprises multiple CNN combined to form a single model (Figure 3.). The output is calculated by averaging the outputs of the numerous CNN. This is a relatively new model, and we will be implementing five-channel MC-CNN (Figure 4.). Previous research has found that the overall accuracy is directly attributed to the number of channels used in the models. However, it's limited only to a certain extent.

The proposed MC-CNN outperforms CNN in many aspects such as:

1) Architecture

MC-CNN is superior in architectural complexity and CNN's ability since many CNN columns combine to form a single MC-CNN Network. The Cerebral Cortex inspires this architecture in the brain, whose structure combines several micro-columns of neurons.

2) Selection of the number of channels

From the previous research works, we observed that accuracy and the number of channels are directly proportional. However, this is limited to the extent of 9 channels, after which the increase in accuracy becomes negligible.

3) Use of Automatic Deep learning

This improves the coordination between all the layers by helping the model adapt to every single deep learning layer. Subsequently, this increases the classification accuracy manifold. Therefore, it eases the image's overall classification process.

4) The dual advantage of the model

The individual CNNs are sufficiently enough to produce fair results by setting new, improved accuracy during training and testing. This fine structure of CNN is advantageous for us since many CNNs combine to form a single MC-CNN. Hence the results will be enhanced multifold.

5) Training and Testing of Data

Our proposed five-channel MC-CNN consists of several layers and is created by combining five individual CNNs. [12] various pre-processing techniques are used for enhancing images. Here P0 refers to original images, whereas P1 relates to images that underwent contrast enhancement. The next step would be image augmentation using various transformations such as cropping and flipping. Here we have represented the normal image as D0, flipped image as D1, and cropped image as D2.

We have tested the trained dataset using five individual CNNs for performing the multi-class weapon detection. Here four labels are assigned for four different classes of guns. The dataset is run through the five CNNs. Finally, we obtained the results by taking an average of all the individual values. The formulae for the same can be written as seen in Eq. 2.

$$y^{i}MC - CNN = \frac{1}{5}\sum_{j}^{streams} y^{i}\,CNN_{j} \qquad (1)$$

Here j denotes the j^{th} channel of the five CNNs.

Algorithm

1. Retrieve the original dataset.
2. Perform Contrast Enhancement to enhance the images in the dataset.
3. Augment dataset and create two more datasets using flipping and cropping of original images.
4. Resize all the images to 128 x 128 pixels.
5. Set the solver modes as CPU.
6. Set initial learning rate base-lr as 0.001 and start training the model.
7. Set the test batch volume batch as 246 for the test, test batch test-iter as 6, and test interval test-interval as 250.
8. Test for every 250 iterations, calculate the average of all the seven CNNs, and display the classification accuracy with time.

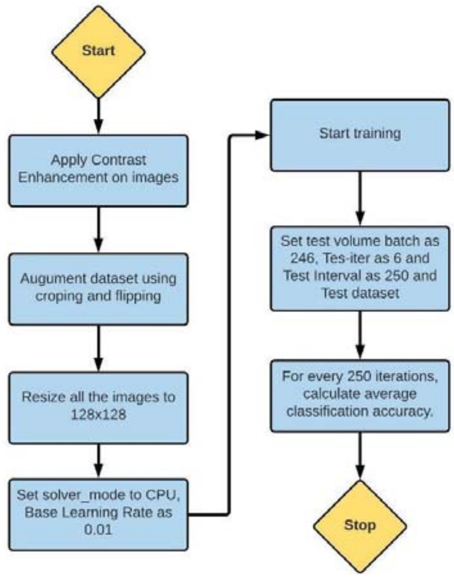

Fig. 3. Flowchart of MC-CNN

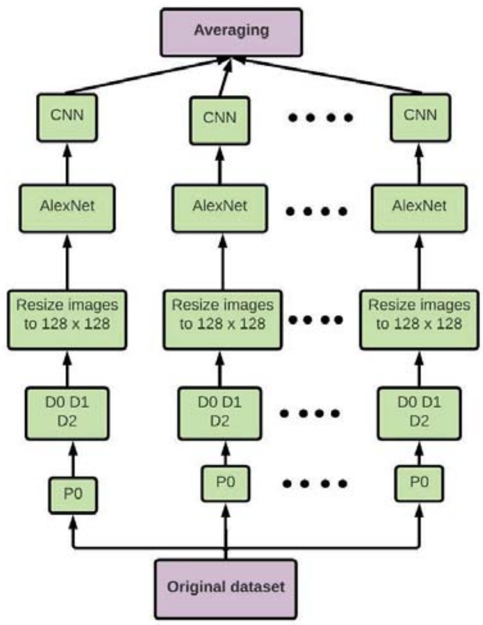

Fig. 4. Architecture of MC-CNN

B. Faster Region-Based Convolutional Neural Network

Faster Region-Based Convolutional Neural Network (Faster R-CNN) is a deep Convolutional Neural Network that appears as a single, end-to-end, unified network used in object detection applications.

The first layer of Faster R-CNN deals with RPN (Regional Proposal Network) to find the foreground class. In the layers, the foreground is somewhat divided Into anchor boxes. After several layers, it finds IOU (Intersection Over Union) {also known as overlapped area} region, to give faster and accurate results. (Figure 6.)

Feature maps of the foreground class are then sent as input to the subsequent layers for ROI pulling. ROI task is to set up a uniform size for all the input feature maps. The whole process of R-CNN will buckle in a few sections, also yielding the brightest results.

Faster R-CNN solved the slow processing issue of the selective search approach, which is used in R-CNN. To extract the objects from images, Faster R-CNN uses the Regions Proposal Network (RPN) with various scales and aspect ratios, easing the image's object extraction.

Faster R-CNN introduced the concepts of anchor boxes. Each anchor box is associated with a specific scale and aspect ratio. It consists of two modules:

RPN: to generate region proposals, each proposal has its probability score of being an object and its label.

Fast R-CNN: It consists of an ROI pooling layer. It takes the previously generated proposals as input, allowing the reuse of the feature map from CNN and significantly speeds up the train and test time.

Algorithm

1) Retrieve the original dataset.
2) Perform Contrast Enhancement to enhance the images in the dataset.
3) Augment dataset and create two more datasets using flipping and cropping of original images.
4) RPN generates Region Proposals.
5) For all region proposals in the image, a fixed-length feature vector is extracted from each region using the (Region of Interest) ROI Pooling layer.
6) The extracted feature vectors are then classified using the Fast R-CNN.
7) The classification accuracy of the detected objects in addition to their bounding-boxes is displayed.

The algorithm can be viewed pictorially in Figure 5.

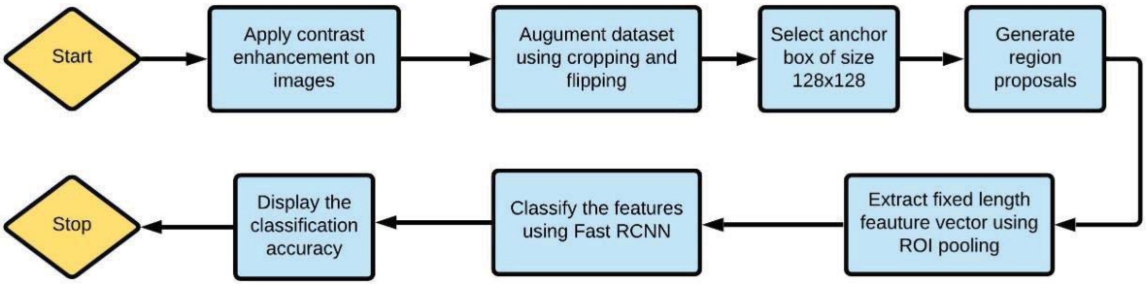

Fig. 5. Flowchart of Faster R-CNN

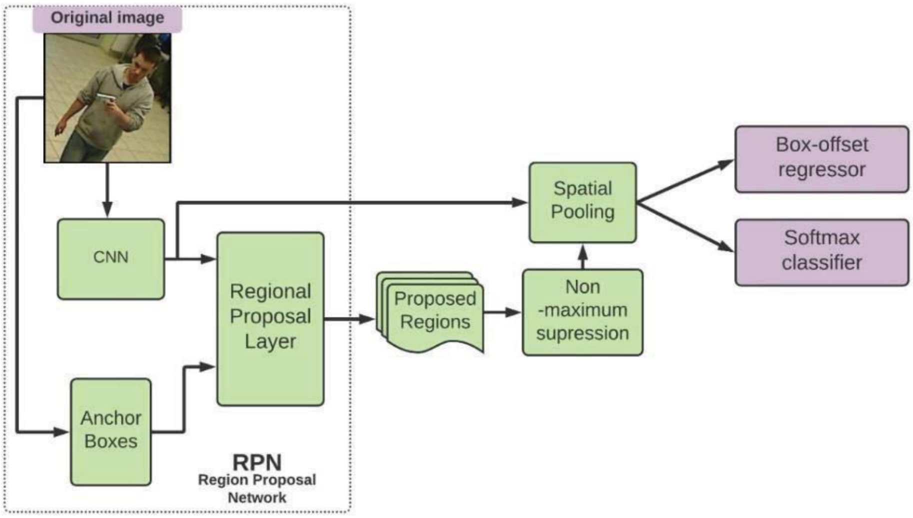

Fig. 6. Architecture of Faster R-CNN

V. Results and Discussion

A. Experimental Process

We have used a custom-built dataset by selecting the most technical and appropriate images from google to build an efficient model. As seen in Table 1, all gun types are equally distributed in training the dataset. This ensures that there is no bias in calculating the accuracy of the models used. All the dataset images used are 3 dimensional (length, width, channels) RGB, where there are 3 channels for RGB(Red, Green, Blue channels) which means that all the dataset images used are color images. Our dataset images are minimally pre-processed to make sure that the images don't lose their originality. This helps in extracting all the features by the proposed models to make accurate weapon prediction. For the experimentation, we have used an Intel i7 (8^{th} generation) processor – max clock speed of 2.6 GHz – with a turbo speed up to 4.5 GHz, integrated with Intel integrated UHD graphics, and a four-core NVIDIA GeForce GTX 1650 Max-Q dedicated graphic card. PyTorch has been used because of its ease of use in utilizing the GPU for parallel processing and other libraries used to load and make batches of images before being trained by the deep learning model. The batch size is set to 24 since it yielded the best possible results.

TABLE II. EXECUTIONAL RESULTS

Testing Parameters	MC-CNN	Faster R-CNN
Accuracy (%)	94.259	92.573
Validation Accuracy (%)	93.174	90.013
Sensitivity (%)	97.277	96.461
Specificity (%)	93.334	94.746
Number of layers	5	5
Training Dataset Size	788	788
Batch Size	24	24
Dropout Ratio	0.6	0.5

B. Interpretation

There are various comparisons between objects and analysis, but in general, a method must opt where the yield provides better results. Likewise, we have various CNN algorithm methods on this topic, out of which MC-CNN throws better results than Faster R-CNN, respectively. We can see from Table II, the accuracy yields to 94.259 and 92.573 for MC-CNN and R-CNN, respectively, where we can see clearly how our method overthrows others despite having a complex structure to interpret or analyse or using this method increases the cost of the process, we can still opt for MC-CNN. Thus, every method today keeps evolving to maintain its visibility in gathering data, selecting the weapon for detection and involving image processing, and comparing the objects to produce better results. Even after studying few cases where Faster R-CNN does not require to be fed with 2000 proposals or more, they lag in performing under a single framework since the architecture is quite simple to MC-CNN. Extraction and classifications happen simultaneously under one frame, adding up to give better results.

C. Discussions

Multi Contrast Convolutional Neural Network, also known as MC-CNN, is formed by clubbing together multiple CNN streams. The average of the outputs of all the CNN combined yields the MC-CNN output. This provides the best results when compared to all the CNN streams kept linearly. The main architectural difference between MC-CNN and Faster R-CNN is that in Faster R-CNN, the original data set is directly fed into the CNN network. In contrast, for MC-CNN, the original data set is sent to different filters (also termed pre-processors), and the output of the pre-processors is being fed as the input for the CNN streams. These are combined to form an MC-CNN stream network. The filters used for data pre-processors have multiple ranges (For instance, it can have 3, 5, 9 filters). The complexity and results vary based on the number of filters used.

VI. Conclusion

Real-time detection and the dynamic classification of concealed hand-held weapons thorugh analysis of surveillance videos are challenging tasks. Contemporary methods incorporated in the industry for weapon detection utilize such models capable of performing only binary classification. These models are highly inefficient and often produce inaccurate results. Our paper presents state-of-the-art frameworks and models capable of accurately and quickly detecting concealed weapons along with them classifying them into appropriate types. We created our own expansive data set consisting of 788 images to train and test our model for this work. Furthermore, we have performed a detailed and comprehensive comparative analysis of the two algorithms based on various relevant factors such as accuracy, specificity, sensitivity and execution time. It was observed that the utilization of our original data set without pre-processing couldn't yield satisfactory results. However, consolidating them in MC-CNN gave us better outcomes. The experimental results reveal that MC-CNN has a more complex architecture than Faster R-CNN and yields far better results. The proposed MC-CNN algorithm outperforms the Faster R-CNN with an accuracy of 94.259%. The proposed work has immense scope for utilization in different industries. Our model falls short to yield results with incredible exactness due to a limited number of images in the testing and training data set. In the future, we plan to run the algorithm in industrial real-time scenarios. Our proposed work demonstrates that with an improved data set and more versatile strategies, better results can be achieved.

Acknowledgment

We want to thank Prajval Mohan and Lakshya Sharma from the School of Computer Science and Engineering at Vellore Institute of Technology, Vellore, India, for their expertise and guidance throughout our project.

References

[1] Tiwari, R. K., & Verma, G. K. (2015). A computer vision based framework for visual gun detection using harris interest point detector. Procedia Computer Science, 54, 703-712.

[2] Lim, J., Al Jobayer, M. I., Baskaran, V. M., Lim, J. M., See, J., & Wong, K. (2021). Deep multi-level feature pyramids: Application for non-canonical firearm detection in video surveillance. Engineering Applications of Artificial Intelligence, 97, 104094.

[3] Jain, D. K. (2019). An evaluation of deep learning based object detection strategies for threat object detection in baggage security imagery. pattern recognition letters, 120, 112-119.

[4] Lai, J., & Maples, S. (2017). Developing a Real-Time Gun Detection Classifier. Course: CS231n, Stanford University.

[5] Verma, G. K., & Dhillon, A. (2017, November). A hand-held gun detection using faster r-cnn deep learning. In Proceedings of the 7th International Conference on Computer and Communication Technology (pp. 84-88).

[6] Pérez-Hernández, F., Tabik, S., Lamas, A., Olmos, R., Fujita, H., & Herrera, F. (2020). Object detection binary classifiers methodology based on deep learning to identify small objects handled similarly: Application in video surveillance. Knowledge-Based Systems, 194, 105590.

[7] González, J. L. S., Zaccaro, C., Álvarez-García, J. A., Morillo, L. M. S., & Caparrini, F. S. (2020). Real-time gun detection in CCTV: An open problem. Neural networks, 132, 297-308.

[8] Jain, H., Vikram, A., Kashyap, A., & Jain, A. (2020, July). 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). IEEE.

[9] Olmos, R., Tabik, S., & Herrera, F. (2018). Automatic handgun detection alarm in videos using deep learning. Neurocomputing, 275, 66-72.

[10] Thakare, A. (2018). Concealed Weapon Detection Using Image Processing. International Journal of Electronics, Communication and Soft Computing Science & Engineering (IJECSCSE), 31-34.

[11] Grega, M., Matiolański, A., Guzik, P., & Leszczuk, M. (2016). Automated detection of firearms and knives in a CCTV image. Sensors, 16(1), 47.

[12] Mohan, P., Pahinkar, A., Karajgi, A., Kumar, L. D., Kasera, R., Gupta, A. K., & Narayanan, S. J. (2020, November). Multi-Contrast Convolution Neural Network and Fast Feature Embedding for Multi-Class Tyre Defect Detection. In 2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA) (pp. 1397-1405). IEEE.