Multimedia Systems (2022) 28:347–360
https://doi.org/10.1007/s00530-021-00848-9

ORIGINAL RESEARCH PAPER

Employing data generation for visual weapon identification using Convolutional Neural Networks

Neelam Dwivedi[1] · **Dushyant Kumar Singh**[1] · **Dharmender Singh Kushwaha**[1]

Received: 24 February 2021 / Accepted: 17 September 2021 / Published online: 1 October 2021

Abstract

The use of weapons nowadays is becoming a leading cause of severe crimes in our society, which reluctantly results in dreadful consequences. The weapons used typically varies from knife, iron-rod, dagger, sabre to firearms like guns and bombs. Due to the unavailability of any proactive mechanism for avoiding heinous crimes using such weapons, an active surveillance performing real time weapon identification is proposed here, as a boon to societal security requirement. As part of it, this paper presents a novel approach based on Convolutional Neural Network (CNN) for identifying visual weapons. This proposed CNN model is initialized with the pre-trained Visual Geometry Group-16 (VGG-16) network weights. These weights are further fine-tuned by training this CNN model with comprehensive weapons (knives and handguns only) and non-weapon images. Weapon category images correspond to further classified classes of "isolated" and "handheld" weapons. However, weapon identification is challenging because of unavailability of diverse databases containing images with variations in shape, texture, scale, occlusion of weapon, etc. This paper reduces this limitation by presenting an algorithm for generating new images and other algorithm for preprocessing the images for quality enhancement. The accuracy achieved is 98.07% with original isolated images and 98.36% with its preprocessed images, while 98.42% with original handheld images and 98.80% with its preprocessed images. The preprocessed algorithm's applicability is confirmed by the higher accuracy achieved by this model using preprocessed images. The accuracy achieved is on an average of ~7% higher than those achieved by other researchers with similar work. The improved result of weapon identification in terms of accuracy proves the appropriateness of the proposed research in being used commercially.

Keywords Weapon classification · Transfer learning · Convolutional Neural Network · VGG-16

1 Introduction

An individual carrying firearms in public places is a strong indicator of conceivable hazardous circumstances. These incidents are increasing nowadays in which individuals or small groups make use of small weapons such as handguns and knives to injure or kill as many people as possible. According to National Crime Records Bureau statistics, a total of 17,490 murders were committed using guns in India during 2010–2014 [1]. The absolute number of firearm deaths went up to 26,500 in 2016 [2]. Robberies, snatching, looting etc., are the major crimes that are happening in public places by showing guns, knives etc. To prevent these crimes government is installing CCTV cameras in public places. Video footage obtained through these CCTV cameras is regularly monitored by the security personnel. Success rates of timely detection of crime happening depend on the attention of the operator. A person can not recognize the activities correctly after a gap of 30-40 minutes, if monitoring the video stream continuously as per the reports published by Velastin et al. [3] in the year 2006. Another report published in Security Oz Magazine [4] claims that the miss rate of recognizing such activities in video footages obtained through CCTV camera increase up to 95 % after 22 min. From these studies, it can be concluded that observing all the video feeds obtained through multiple CCTV cameras on a single screen attentively is not possible for a human

✉ Neelam Dwivedi
neelamdw@gmail.com

Dushyant Kumar Singh
dushyant@mnnit.ac.in

Dharmender Singh Kushwaha
dsk@mnnit.ac.in

[1] Motilal Nehru National Institute of Technology Allahabad, Prayagraj, India

being. Thus, installing CCTV cameras is not sufficient for recognizing the crime effectively. To overcome these limitations, automatic video surveillance equipped with an accurate automatic weapon detection alert system is of critical requirement.

A need for such system increases in public events where a large number of people gathering is anticipated. Mahakumbh Mela in India is such an event where lakhs of people participate. In 2019, Juna Akhara saint complained of life threaten by three armed person, to the Mela police [9]. It happens despite of the ban on any kind of weapon and even the presence of CCTV cameras based surveillance. Any active weapon identification system in such circumstances might have helped avoid carrying of deadly weapons in public. This will also help in alerting the security personnel as soon an attempt is made for such inhuman activities. So, automatic identification of weapons can improve the efficiency of video surveillance [5, 6], smart homes [7], detection of security breaches in smart cities [8], and other application domains. The nature and extent of crime can also be predicted by classifying the types of identified weapons. This will help in reducing the losses by taking appropriate preventive actions on time. It is reported that a maximum number of crimes, such as looting, murder, snatching, etc., are committed either using a gun or knife as these are easy to carry. Due to these reasons, the identification of two classes of weapons: knife & handgun, is approached in this work.

Weapons may be segmented, detected, and classified to train the model using a traditional machine learning approach where required features are selected manually. A deep learning approach can also be used where desired features are extracted from image by the model itself. Deep learning is a successful approach for video summarization [10–12], video skimming [13], image analysis because it defines a learning algorithm that uses the characteristics of pattern, points, edges, etc., and analyses the images via color distribution. Therefore, a deep learning-based approach is used for the visual weapon identification in this work. Convolutional Neural Network (CNN) is one of the most popular deep learning approach having multiple trainable layers which can be trained robustly for many types of problems. It consists of convolutional layers, pooling layers and fully connected layers. Convolutional layers extract the features which are later used by fully connected layers for learning. Pooling layers progressively reduce the spatial size of the representation of the feature to reduce the amount of parameters and computation in the network. Pooling layer helps creating robust features to withstand variations in the position of input data. Deep CNNs automatically discover increasingly abstract features from raw data [14, 15] which makes it suitable for various applications such as object segmentation, detection, and classification.

A CNN model requires large amount of raw data and high speed computational machine such as GPUs for the accurate and faster training of the model. These constraints can be minimized using the transfer learning technique where weights of pre-trained model of similar applications are used for initializing the weights of convolutional layer of a new CNN model. Initial layers of the CNN extract low level features such as edges, corners while deeper layers of the CNN extract higher level features that are specific to the application. Thus, weight of deeper layers should be fine-tuned with the data specific to the application. Fine-tuning the weights of deeper layers require less data as compared to training a new CNN model from scratch. Still there are many challenges that persist in automatic identification of visual weapons. These are as mentioned below:

- Unavailability of number of diverse images of guns and knives
- Quality of the available images of guns and knives
- Guns and knives are handled in different ways using one or two hands, due to this large part of the weapons may be occluded.

Generative Adversarial Networks (GANs) as an approach which can be used to generate new images looking over the need of a diverse and more exhaustive dataset. The model must first be trained with enough images before it can create new images using GANs. We devised an approach to augment the photos instead of utilizing GANs because several acceptable images were not available. We have proposed an algorithm to augment the images instead of using GANs because several adequate images were not available. The proposed algorithm in this paper generates images similar to the input image, which have different shapes, textures, scales, and orientation. This reduces the challenges which occurs due to unavailability of images. An algorithm for enhancing the quality of images to overcome the challenges of image quality is also presented here. VGG-16 based a novel deep CNN model has also been proposed here for the visual weapon identification objective. VGG-16 is a deep CNN model trained on ImageNet database having 1.28 million images approx. of over 1000 object classes. Thus, the weights of newly proposed models for visual weapon identification are initialized with the weights of pre-trained VGG-16 model. Thereafter, weights of proposed model in this work are fine-tuned using the images of handguns, knives and non-weapon classes. Following are the major contribution of this work:

- A new algorithm for generating images similar to input images; and having different shapes, sizes and orientation
- A new algorithm for enhancing the quality of images
- A CNN model for weapon identification

- Analyzes the effect of dropout rate and number of neurons in convolutional layers of CNN for weapon identification

This paper is organized as follows: In Sect. 2, literature survey for weapon detection and classification is presented. Creation of extended database is presented in Sect. 3. In Sect. 4, methodology and architecture of the proposed approach are explained. Experimental results and their analysis are presented in Sect. 5. Conclusion is presented in Sect. 6.

2 Literature survey for weapon identification

In this section, a brief review of the state-of-the-art approaches of weapon identification and classification is presented. Weapon identification and classification methods can be broadly classified into two categories: traditional methods and deep learning methods. Fuzzy classification, Active Appearance model, Harris corner detector, Support Vector Machine, Random Forest, Bag of Visual Word are some of the popular traditional methods used for weapon identification and classification. Convolutional Neural Network (CNN), Recurrent Convolutional Neural Network (R-CNN), Faster R-CNN, Overfeat Network are deep learning methods used by researchers to identify and classify the weapons.

Many researchers are working in the area of weapon detection and classification since long. Maksimova et al. [17] present a classification model for knife detection. In this model, fuzzy clustering approach is used to detect the knife present in the frames of a video. Glowacz et al. [18] present a knife detection method in the images. This method is based on Active Appearance Models [19] and Harris corner detector [20]. Interest points are designated by the Harris corner detector, which is used to detect the knife tip. Thereafter, Active Appearance Models of knife is created using this tip. The overall performance of this architecture depends on Harris corner detector accuracy. Kmiec et al. [21] presented an Active Appearance Models to detect the presence of a knife in an image. Accuracy of their proposed approach for detecting the knife is ≈ 93%. To improve the efficiency of the knife detection in video surveillance, Kmiec et al. [22] proposed a new image feature, named as "Dominant Edge Directions (DED)". This new feature is examined by training a Support Vector Machine (SVM) classifier with a linear kernel. The proposed DED feature as input for detecting the knife in a video. They claimed that the detection speed is improved using the DED feature but accuracy is not much improved. Kmiec et al. [23] further proposed an approach to detect the images of knives in different backgrounds having varying lighting conditions using SVM classifier, trained with Histograms of Oriented Gradients (HOG) feature descriptor. Buckchash and Raman [24] presented a new object detection algorithm for detecting and classifying a visual knife. This algorithm uses Gaussian mixture for detecting the foreground object, key point detectors for localizing the object and Multi-Resolution Analysis for classifying the object. Many researchers are also working to detect the firearms such as handguns, pistols etc., to further improve the automatic surveillance system. Tiwari and Verma [15] presented automatic surveillance system which has the capability of visual gun detection. The system uses color based segmentation and Harris interest point detector to detect the gun in an image. The proposed approach takes long time to detect a gun in an image thus not suitable for real time application. Performance of their approach is also not appropriate with the videos having variations in the ambient illumination. Grega et al. [25] proposed an algorithm for detecting firearm and knife in an image followed by the alert generation.

Many researchers [14, 27] have also exercised some deep learning algorithms to detect and classify the weapons in a video. Susarla et al. [26] presented a Structural Recurrent Neural Networks (SRNN) model to recognize the spatio-temporal human-object interactions in video surveillance. Lai and Maples [27] developed a real time gun detector using a Tensorflow-based implementation of the Overfeat network. Performance of their proposed approach was 93% for training database and 89% for testing database after tuning the hyper parameters. Olmos et al. [14] presented a pistol detection system in a video based on faster R-CNN. The authors created their own dataset for the training and testing purpose. Further, they reformulate their detection problem to minimize false positive rate. Verma and Dhillon [28] presented a handheld gun detection using faster R-CNN deep learning approach. They achieved 93% approx. accuracy for their proposed work. Egiazarov et al. [29] presented a weapon detection system using an ensemble of semantic CNNs. They claimed that their proposed approach decomposes the problem of detecting and locating a weapon into a set of smaller problems, concerned with the individual component parts of a weapon.

All of these existing approaches suffer from low accuracy and detect only one category of weapons. There is a need of an approach that can detect and classify more categories of weapons. Thus to improve the accuracy, new architecture has been proposed in this paper and also analyze the effect of increase or decrease of dropout rate and number of neurons in the network. In addition to this, the proposed architecture has been trained on the similar images with variations in shape, size and orientation to make the model invariant to the shape and size. Development of database has been discussed in the next section.

3 Database development: generation of new images similar to input image followed by quality enhancement

This section discusses the development of a large size and robust database of visual weapon, by generating number of images similar to input image, but having variations in shape, size and orientation.

For this work, images of two weapon classes: knife & handgun, and one non-weapon class are used. Non-weapon class contains the images of the objects such as wrenches, screws, mobile phones and many more. These objects are selected for non-weapon class because some of them are frequently being seen in the public places and can be used as a weapon if required. Some images are downloaded from the Internet and many others are synthesized in the laboratory by author herself. Input images have been further categorized into two categories: isolated images and handheld images. The collection of isolated images is named as Database #1 while collection of handheld images is named as Database #2. Database #1 is identified as a low-risk because objects in the images are isolated. Database #2 is identified as a high-risk because objects in the images are handheld that can be used instantly. Figures 1 and 2 present the sample images of Database #1 and Database #2 respectively.

Total of 624 isolated images and 808 handheld images each of knife, handgun and non-weapon class have been acquired. Performance of any CNN model highly depends on the amount and quality of the data being used for training. Therefore, to increase the number of images and diversify the database, an algorithm is proposed for generation of new images similar to input image but having variations in shapes, textures, scales and orientation. The proposed Algorithm 1 is used for generating *k* new images similar to the input image. New images are generated using this three step procedure. These steps are:

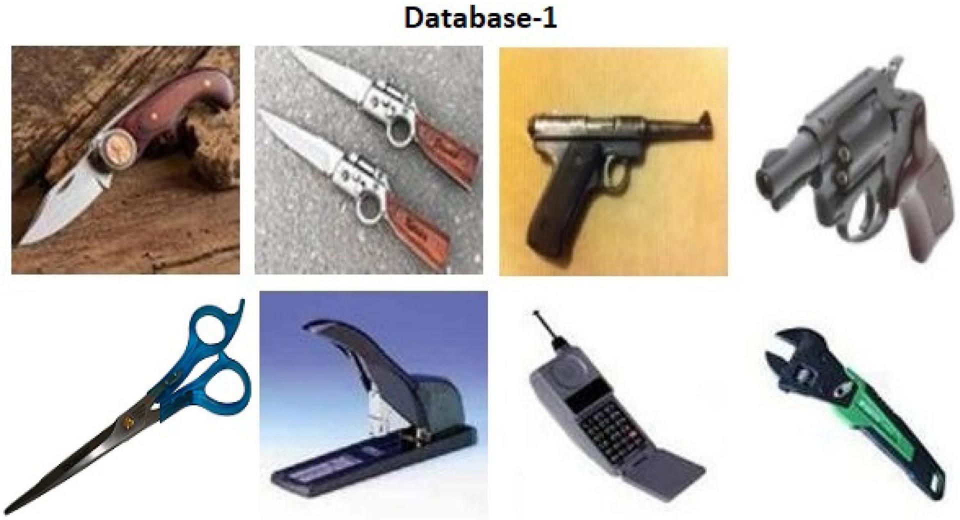

Fig. 1 Sample images of Database #1: Top row contains images of isolated weapons (knife & handgun) while bottom row contains isolated images of non-weapon class

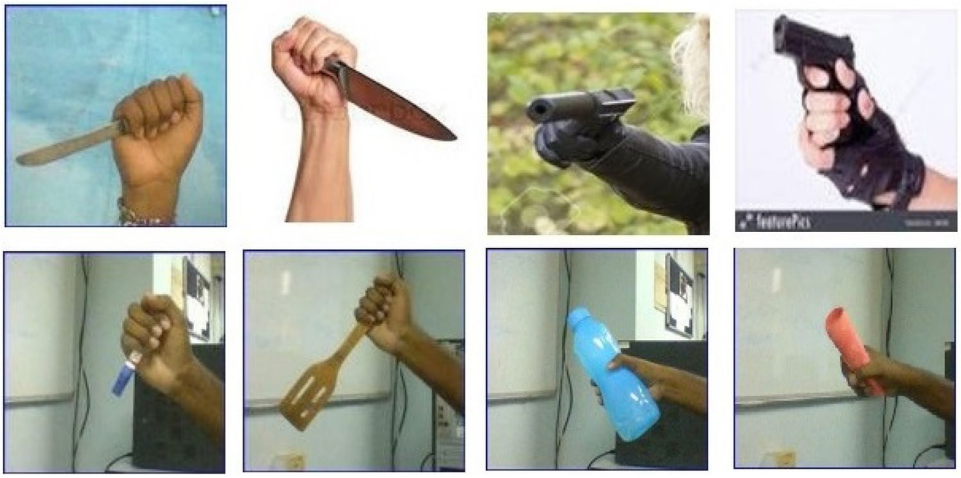

Fig. 2 Sample images of Database #2: Top row contains images of handheld weapons (knife and handgun) while bottom row contains handheld images of non-weapon class

1. Two dimensional input image is transformed into 3 dimensional space.
2. Image is rotated with x, y and z axes by angles $\theta 1, \theta 2$ and $\theta 3$ respectively.
3. Rotated image obtained from step 2 is projected in the two dimensional plane. Thus, obtained image is similar to the original image having variations in shapes, scales and orientation.

After generation of new images, there is need of quality enhancement for some of the poor quality images. In this paper, a new Fuzzy based technique is proposed to enhance the quality of the images by contrast stretching.

Detail procedures of generation of new images and their quality enhancement are explained in Sects. 3.1 and 3.2 respectively.

3.1 Generation of *k* new images similar to an input image

The images of weapons class (Knife & Handgun) and non-weapon class (Wrenches, screws etc.) are acquired by different medium as explained earlier. Although the variety of these images gathered is insufficient for accurate model training, a new Algorithm is presented to generate new images, as previously mentioned.

This generates n number of images from an input image in one iteration. Value of n depends on number of boundary types used in an iteration. To generate new images, a 3-D transformation matrix is derived first followed by the estimation of a projection matrix corresponding to different boundary type B_t. For this, a scaling factor s is calculated by multiplying the maximum of width and height of the original image and a constant c ($c = 2$, for this work). Thereafter, three random angles $\theta 1,\ \theta 2$, and $\theta 3$ are determined. Shapes and orientation of new generated images will depend on the values of these angles $\theta 1,\ \theta 2$, and $\theta 3$. R^x, R^y and R^z are following three rotation matrix along X- axis (R^x), Y-axis (R^y) and Z-axis (R^z) respectively.

$$R^x = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos(\theta 1) & -\sin(\theta 1) & 0 \\ 0 & \sin(\theta 1) & \cos(\theta 1) & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

$$R^y = \begin{bmatrix} \cos(\theta 2) & 0 & \sin(\theta 2) & 0 \\ 0 & 1 & 0 & 0 \\ -\sin(\theta 2) & 0 & \cos(\theta 2) & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

$$R^z = \begin{bmatrix} \cos(\theta 3) & -\sin(\theta 3) & 0 & 0 \\ \sin(\theta 3) & \cos(\theta 3) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

Algorithm 1 Generation of similar images as of input image

```
INPUT: Image $I$ of size $h \times w$, boundary type $B_t$.
OUTPUT: Image $(I_N)$ = N number of images.
Procedure $ImageGeneration(I, B_t)$
    Scale factor $(s)$ = maximun (h: height of the input image $(I)$, w: width of the input image $(I)$) $\times$ constant $(c)$.
    Determines three random angles $\theta_1$, $\theta_2$ and $\theta_3$.
    New rotation matrix $(R)$ = rotation matrix $(R^x)$ rotated along $X - axis$ with angle $\theta_1$ $\times$ rotation matrix $(R^y)$ rotated along $Y - axis$ with angle $\theta_2$ $\times$ rotation matrix $(R^z)$ rotated along $Z - axis$ with angle $\theta_3$ .
    Determines four homogeneous coordinates $(s_1, s_2, s_3$ & $s_4)$ of source matrix $(s)$ using four corner of the input image $(I)$.    ▷ Assuming origin coincides with centre of the input image $(I)$.
    Rotated transpose matrix $R^T$ = rotated matrix $(R)$ $\times$ transpose of the source matrix $(s^T)$.
    $s_z = translate$ rotated transpose matrix $R^T$ by $s$ unit towards $Z - axis$.
    Transformation matrix $(TM) = (R^T + s_z) \times s$.
    Projection matrix $(P)$ = taking a $projection$ of $TM$ into 2-D plane to a given boundary type $(B_t)$.
    New image $I_{new} = I \times P$
    return$(I_{new})$
end procedure
```

A new rotation matrix (R) is obtained by multiplying R^x, R^y and R^z. Thus, R is a rotation matrix along X, Y, and Z axes with angles θ1, θ2 and θ3, respectively.

$$R = R^x \times R^y \times R^z$$

Matrix R is multiplied with the homogeneous coordinates of source image to generate new coordinate of the image. Homogeneous coordinates of source image is depicted by assuming that origin coincides with the center of the image. Thus, homogeneous coordinates of source image is:

$$\text{source} = \begin{bmatrix} -\frac{w}{2} & \frac{h}{2} & 0 & 1 \\ \frac{w}{2} & \frac{h}{2} & 0 & 1 \\ \frac{w}{2} & -\frac{h}{2} & 0 & 1 \\ -\frac{w}{2} & -\frac{h}{2} & 0 & 1 \end{bmatrix}$$

where, w and h are the respective width and height of the source image.

Now, by multiplying rotation matrix R and transpose of source matrix (source^{T}), a new matrix R_{new} ($R_{\text{new}} = R \times \text{source}^{\text{T}}$) is generated. R_{new} is a rotated source matrix along X, Y, and Z axes with angles $\theta 1$, $\theta 2$, and $\theta 3$ respectively. This converts the source image from 2-D plane to 3-D surface but center of the image coincides with origin. Thereafter, this matrix R_{new} is translated by s unit towards Z-axis (s_z) which changes its position. A new transformation matrix TM is obtained by multiplying the translated matrix with the source matrix.

$$\text{TM} = (R_{\text{new}} + s_z) \times \text{source}$$

Next, a projection matrix P is obtained by taking the projection of TM into 2D-plane to a given boundary type B_t. Now, by multiplying this projection matrix P with original input image I, a new image I_{new} is generated. Two types of boundaries (B_t: reflection and edge) have been used in this work. Algorithm 1 shows the complete procedures for generation of new images similar to input image I. By applying the Algorithm 1; 1872 isolated images and 2424 handheld images each of knives, handguns and non-weapons are added to the Database #1 and Database #2 respectively.

Table 1 Summary of images of Database #1 and Database #2

		Total #image	Actual #image	Generated #image
Database #1	Train	4680	1560	3120
	Test	936	312	624
Database #2	Train	6024	2008	4016
	Test	1224	408	816

Figure 3b and c present new generated images by applying Algorithm 1 on images of Fig. 3a with boundary type: 'edge' and 'reflection' respectively. Visuals of images of Fig. 3a–c confirm that the new images generated by applying Algorithm 1 is similar to the original images but having variations in shapes, textures, scales and orientation. By changing the rotation angles $\theta 1$, $\theta 2$ and $\theta 3$, new rotation matrix R can be generated. New images can be generated corresponding to every new rotation matrix.

Table 1 summarizes the total number of images, actual images and generated images with Algorithm 1 in training set and testing set for Database #1 and Database #2 respectively.

Once, all the images are generated then preprocessing algorithm is applied to enhance the quality of these images as discussed in the next subsection.

3.2 Preprocessing of input image

Quality enhancement is necessary for the images obtained from the videos of low-resolution CCTV cameras. This section presents the details of quality enhancement algorithm named as preprocessing Algorithm 2.

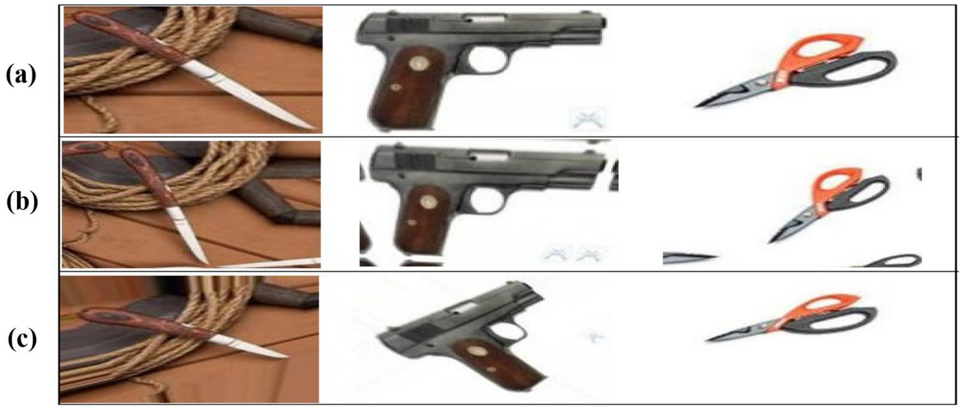

Fig. 3 **a** Original image of knife, handgun and scissors **b**, **c** Generated images from the original image by applying Algorithm 1

Algorithm 2 Preprocessing of input image

```
INPUT: Input image $I$.
OUTPUT: Enhanced image $I_e$.
Ensure:
    (i) Values of variables $k_1$ and $k_2$ depend on input image and lying in between 1 to 255.
    (ii) $I_y^i$ represents intensity value of $i^{th}$ pixel of Y-component of the $YC_bC_r$ image.
Procedure $preprocessedImage(I)$
$I_{YC_bC_r} \leftarrow RGBtoYC_bC_r(I)$          ▷ converting $rgb$ color image into $YC_bC_r$ color image
$I_y \leftarrow ExtY(I_{YC_bC_r})$                ▷ extracting Y component of $YC_bC_r$ color image
$h \leftarrow Hist(I_y)$                          ▷ histogram of $I_y$ image
$M \leftarrow Mean(h)$                            ▷ calculating the mean intensity value of histogram $h$
for each pixel of the image do
    $sp_1^i \leftarrow \frac{I_y^i}{M}$                       ▷ First stretching parameter
    $sp_2^i \leftarrow \frac{255-I_y^i}{255-M}$               ▷ Second stretching parameter
    if $(I_y^i < M)$ then
        $I_e \leftarrow I_y^i + sp_1^i * k_1$
    else
        $I_e \leftarrow I_y^i * sp_2^i + (k_2 - sp_2^i \times k_1)$
    end if
end for
$I_{rgb} \leftarrow$ converting $YC_bC_r$ color image into $rgb$ color image
return($I_e$ )
end procedure
```

This algorithm enhances the quality of an image by stretching the pixel intensities in YC_bC_r color space. YC_bC_r color space is used because brightness or intensity of pixels have confined only in Y component. To stretch the pixel intensity, mean intensity value M of an image I is calculated first by taking the average value of Y component of all the pixels of I. Further, pixels of I is divided into two classes: C1 and C2. C1 class contains pixels having intensity values lying between $[0,\ M-1]$ and C2 class contains pixels having intensity values lying between $[M,\ 255]$. Here, it is assumed that the intensity closer to M should be stretched more while farther from M should stretch less. Thus, separate stretching parameter for each pixel of each class is calculated. Stretching parameter of ith pixel for class C1: $sp1^i$

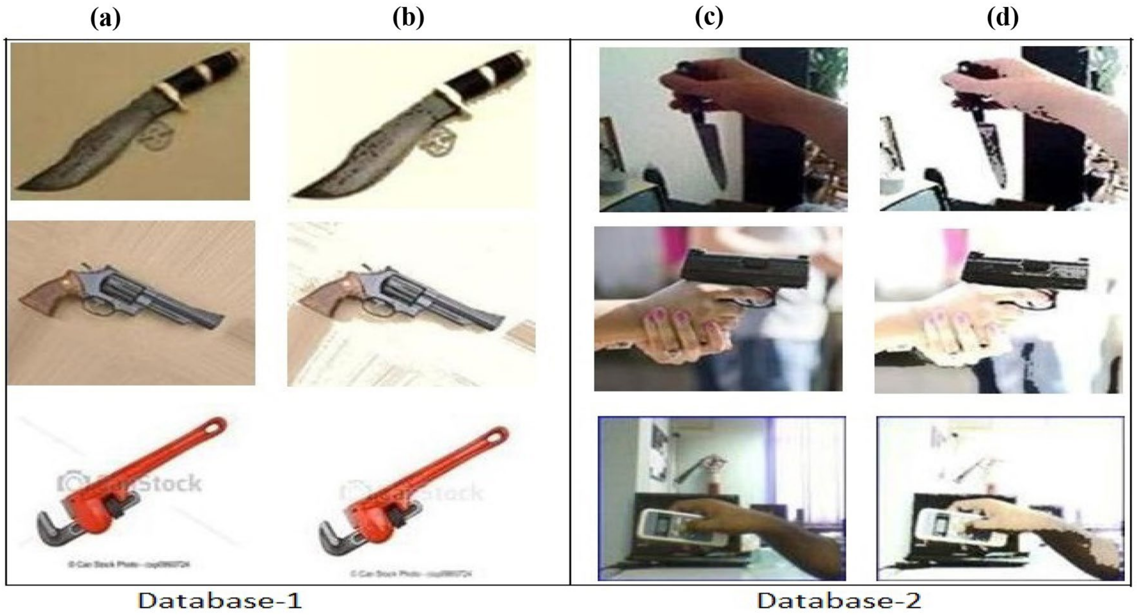

Fig. 4 **a**, **c** Original images from Database #1 and Database #2, **b** , **d** enhanced images obtained by applying Algorithm 2 corresponding to the images of **a** and **c**

Table 2 Details of databases used for the experiments

Database	# of knives images		# of handguns images		# of non-weapon images		Types of images
	Train	Test	Train	Test	Train	Test	
Database #1	1560	312	1560	312	1560	312	Original
Database #2	2016	408	2016	408	2016	408	Original
Database #3	1560	312	1560	312	1560	312	Preprocessed
Database #4	2016	408	2016	408	2016	408	Preprocessed

is defined as the ratio of ith pixel intensity of Y component of the image (I_y^i) and M.

$$sp1^i = \frac{I_y^i}{M}$$

Thus, enhanced ith pixel intensity for class C1 (I_e^i) is calculated as:

$$I_e^i = I^i + sp1^i \times k_1$$

Here, value of constant k_1 (for this work $k_1 = 32$ is selected experimentally) depends on the images in the database. This value need to be adjusted for a database according to the quality of the images.

Separate stretching parameter for each pixel of class C2 element is also calculated by assuming that the intensity closer to 255 should be stretched less whereas farther from 255 should be stretched more. Stretching parameter for ith intensity ($sp2_i$) is calculated as:

$$sp2^i = \frac{k_2 - I^i}{k_2 - M}$$

Thus, enhanced ith pixel intensity for class C2 (I_e^i) is calculated as:

$$I_e^i = I^i \times sp2^i + (k_2 - sp2^i \times k_1)$$

Here, value of constant $k2$ (for this work $k_2 = 255$ is selected experimentally) depends on the images in the database. Once all the pixel intensity are enhanced in YC_bC_r domain, the image is converted back to the RGB color space.

This procedure is applied for every image in the database.

Four databases having different characteristics are developed. Database #1 and Database #2 contain original images while its preprocessed images are kept in Database #3 and Database #4 respectively. Figure 4a presents sample images of Database #1 and (b) presents its respective preprocessed images and stored in Database #3. Figure 4c presents sample images of Database #2 and (d) presents its respective

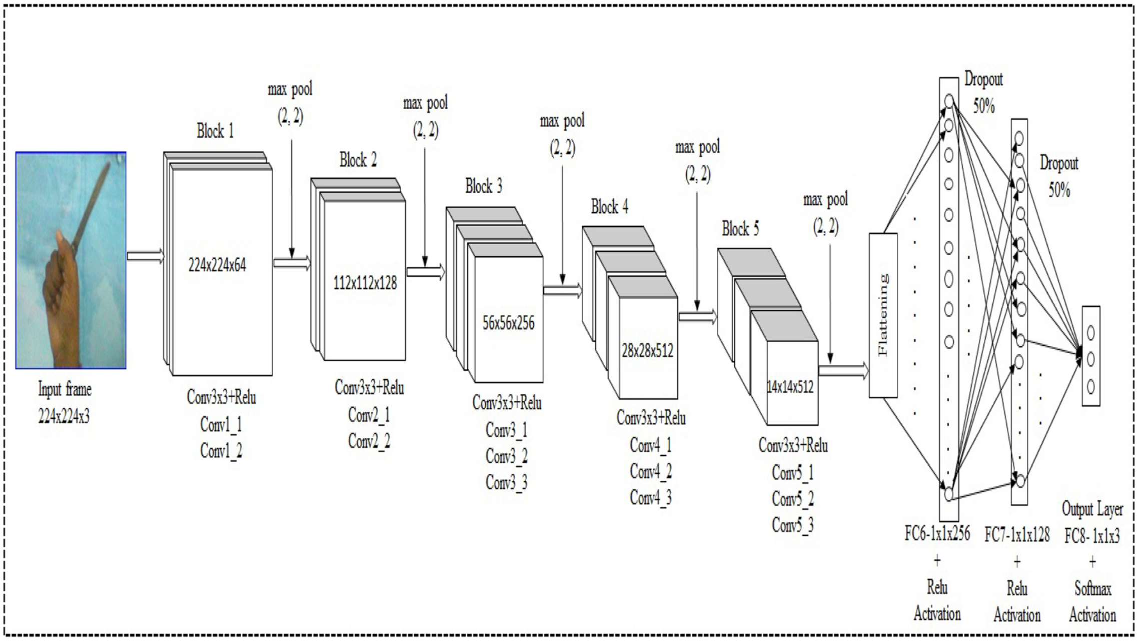

Fig. 5 Architecture of proposed CNN model for visual weapon identification

preprocessed images and stored in Database #4. From this figure it is clear that, after preprocessing visibility of RoI (Region of Interest) in target image is better as compared to the one in original image.

It is also observed that preprocessing algorithm works well even with complex background as this is clearly visible in the case of images of mobile phones and handheld gun.

Table 2 summarizes all details of the databases that are used in this work. Proposed CNN model for visual weapon identification is elaborated in the next section.

4 Proposed CNN model for visual weapon identification

A visual weapon identification using deep CNN architecture is proposed in this paper. The model proposed here is based on the VGG-16 architecture which was proposed by Simonyan and Zisserma [16] for Large Scale Image Recognition. VGG-16 is a 16-weight-layers CNN model which is trained on ImageNet ILSVRC-2014 for classifying 1000 different classes.

Figure 5 illustrates the proposed CNN model being used for visual weapon identification. This model consists of thirteen convolutional layers, five pooling layers, two fully connected layers, and one output layer. These convolutional layers are arranged in five blocks in which first two blocks contain two consecutive convolutional layers each while last three blocks contain three consecutive convolutional layers. Each Convolutional layer of first, second, third, fourth and fifth block consist of 64, 128, 256, 512 and 512 filters of size 3 × 3 respectively. These convolutional layers are used to extract features (horizontal edges, vertical edges, corners, etc.) of an input image. Outer convolutional layer (near to input) extracts low level features such as line, corner of an object while inner convolutional layer (near to FC6 layer) extracts high level features such as object as a single unit. This proposed model contains a max pooling layer after each block of convolutional layers of size 2 × 2. Max pooling layer selects the maximum element from the region of the feature map covered by the filter. Thus, the output after applying the max pooling layer would be a feature map containing the most prominent features of the previous feature map. This model also consists of three fully connected layers (FC6, FC7 and FC8) having 256 neurons, 128 neurons and 3 neurons respectively. Number of experiments are conducted to finalize the number of neurons in fully connected layers FC6 and FC7. These experiments are conducted by increasing and decreasing the number of neurons in FC6 and FC7. Experimentally it is found that this newly proposed model as presented in Figure 5 is having higher classification accuracy as compared to one earlier proposed model by Dwivedi. et al. in their paper “Weapon Classification using Deep Convolutional Neural Network” [30]. Fully connected layers learn from the features extracted by the convolutional layers during the training of the model. Last fully connected layer (FC8) returns the probability of existence of an input image to each class for which the model is trained. Rectified Linear Unit (‘ReLU’) is used as activation function for convolutional layers and first two fully connected layers. It helps in dealing with the vanishing gradient problem [31]. Softmax classifier is used in last fully connected layer (FC8) which calculates the probability of existence of an input image in each target class. 50% dropout rate is applied in FC6 & FC7 which helps to overcome the overfitting problem.

Weight of each parameter of the convolutional layers of this proposed model is initialized with the weights of pretrained VGG-16 model while weights of fully connected layers are randomly initialized. These weights are further finetuned by training the proposed model with the images of all the databases (Database #1, Database #2, Database #3 and Database #4) separately. All the images of these databases have been splitted into 80:20 ratio for training and testing purposes respectively. Once the model is trained, it is tested by evaluating the classification results of the new images.

This section elaborated the proposed CNN model for visual weapon identification in detail. The next section discusses the experimental setup along with the analysis of the obtained results.

5 Experiments and analysis

Classification model is developed and evaluated on python using Anaconda Spyder integrated development environment (IDE). All the experiments are conducted on Intel Core i7 CPU accelerated with 4GB NVIDIA GTX 1050Ti GPU. Total of 4 experiments have been performed to evaluate the performance of the proposed model. Separate experiment is conducted each for Database #1, Database #2, Database #3, and Database #4. Database #1 and Database #2 contain original images while its preprocessed images are stored in Database #3 and Database #4 respectively. Images of Database #1 and Database #3 contain isolated objects which are less harmful while images of Database #2 and Database #4 contain handheld objects which are harmful more likely. Images of all the databases are categorized into three categories: knives, handguns and non-weapons. Size of all the input images of each database is 224× 224 ×3. Output layer predicts the probability of existence of the image for each class. Maximum probability of existence of the image is taken as the output class of that image. Figure 6 presents

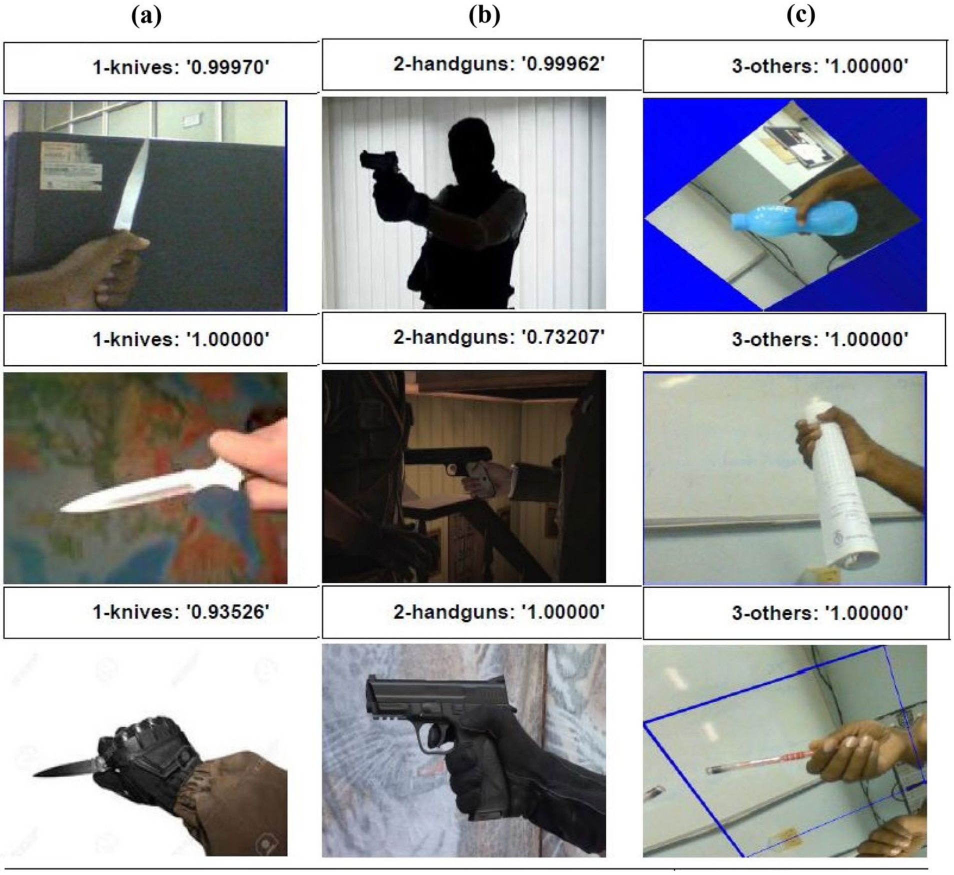

Fig. 6 Output images classified by proposed CNN architecture: **a** presents samples of hand-held knife images, **b** presents samples of handheld handgun images, **c** presents samples of handheld non-weapon images

Table 3 Classification results of a classifier

	Predicted		
Images	Class 1	Class2	Class 3
Class 1 (Actual)	N_1^1	N_1^2	N_1^3
Class 2 (Actual)	N_2^1	N_2^2	N_2^3
Class 3 (Actual)	N_3^1	N_3^2	N_3^3

Table 4 Confusion matrix of Database #1

	Knife	Handgun	Non-weapon
Knife	**1.00**	0	0
Handgun	0.01	**0.98**	0.01
Non-weapon	0.06	0.01	**0.93**

Bold values represent the maximum correctly classified images

samples of output images identified by the trained CNN model for Database #2.

Parameters that are used for measuring the performance of this proposed approach are discussed in the next subsection.

5.1 Parameters used for performance measurement

Precision, recall, F1 score, accuracy and confusion matrix are used to evaluate the performance of this proposed visual weapon identification system. Table 3 presents the classification output for input video frames applied on a classifier, as assumed. Here N_i^j denotes the number of ith activity classified as jth activity by the classifier. Performance metrics for this classification results are calculated as follows:

Table 5 Performance metrics for Database #1

	TP	FP	Precision (%)	Recall (%)	F1 score (%)	Accuracy (%)
Knife	312	21	93.69	100.0	95.68	97.75
Handgun	306	03	99.03	98.07	99.03	99.04
Non-weapon	291	03	98.97	93.27	98.20	97.43

- **Precision**

$$\text{Precision}_i = \frac{N_i^i}{\sum_{j=1}^{4} N_j^i} \tag{1}$$

For correct classification of all activities, the value of precision will be 1.

- **Recall**

$$\text{Recall}_i/\text{Sensitivity}_i = \frac{N_i^i}{\sum_{j=1}^{4} N_i^j} \tag{2}$$

For correct classification of all activities, the value of recall will be 1.

- **F1 Score**

$$\text{F1 score}_i = 2 \times \frac{\text{Precision}_i \times \text{Recall}_i}{\text{Precision}_i + \text{Recall}_i} \tag{3}$$

For correct classification of all activities, the value of $F1$ score will be 1.

- **Accuracy**

$$\text{Accuracy}_i = \frac{N_i^i + \sum_{j,k=1}^{4} N_j^k ; j,k \neq i}{\sum_{j,k=1}^{4} N_j^k} \tag{4}$$

For correct classification of all activities, the value of accuracy score will be 1.

In this application, false-negative has a higher impact as compared to false positive. Thus, while calculating the Fbeta-measure, if needed, higher weightage should be given to recall.

The next subsection presents the experimental results obtained by applying the proposed work on Database #1 which contains images of isolated weapons and non-weapon classes.

Table 6 Confusion matrix of Database #2

	Knife	Handgun	Non-weapon
Knife	**0.98**	0.02	0
Handgun	0.05	**0.95**	0
Non-weapon	0	0	**1.00**

Bold values represent the maximum correctly classified images

Table 8 Confusion matrix of Database #3

	Knife	Handgun	Non-weapon
Knife	**0.98**	0	0.02
Handgun	0.01	**0.98**	0.01
Non-weapon	0.04	0	**0.96**

Bold values represent the maximum correctly classified images

5.2 Experimental results for Database #1

This section analyzes the experimental results obtained for the experiments conducted on Database #1. Tables 4 and 5 presents the confusion matrix and performance metrics respectively. Confusion matrix (Table 4) shows that all the knives are correctly classified while there is some misclassification in handgun and non-weapon classes. From the Table 5, it is found that average precision, average recall, average F1 score and average accuracy obtained are 97.23, 97.11, 97.64 and 98.07 % respectively. Values of all the performance metrics are greater than 97.00% confirming the effectiveness of the proposed CNN model for visual weapon identification.

The next subsection presents the experimental results and their analysis obtained by applying the proposed approach on Database #2 which consists of handheld weapons and non-weapon images.

5.3 Experimental results for Database #2

This section analyzes the experimental results obtained for the experiments conducted on Database #2 containing handheld images of weapons and non-weapon. Tables 6 and 7 present the confusion matrix and performance metrics respectively. Images (Fig. 2 can be referred) of this database consisting diverse background.

From the Table 6, it is observed that all the images of non-weapon class are correctly classified while there is some misclassification in knife and handgun classes.

From the Table 7, it is found that the average precision, average recall, average F1 score and average accuracy obtained are 97.64, 97.63, 98.03, and 98.42% respectively. Values of all the performance metrics higher than 97.50% again confirms the effectiveness of the proposed CNN model for visual weapon identification. By comparing the

Table 7 Performance metrics for Database #2

	TP	FP	Precision (%)	Recall (%)	F1 score (%)	Accuracy (%)
Knife	398	19	95.44	97.54	96.52	97.63
Handgun	389	08	97.98	95.34	97.89	97.79
Non-weapon	408	02	99.51	100.0	99.67	99.84

Table 9 Performance metrics for Database #3

	TP	FP	Precision (%)	Recall (%)	F1 score (%)	Accuracy (%)
Knife	306	15	95.32	98.07	96.52	97.75
Handgun	306	00	100.0	98.07	99.67	99.35
Non-weapon	301	08	97.41	96.47	97.68	97.97

Table 10 Confusion matrix of database #4

	Knife	Handgun	Non-weapon
Knife	**0.98**	0.02	0
Handgun	0.04	**0.96**	0
Non-weapon	0	0	**1.00**

Bold values represent the maximum correctly classified images

experimental results of Database #1 and Database #2, it can be found that performance of this proposed CNN model for weapon identification is comparable for both the databases confirming its robustness.

This subsection presented the experimental results obtained for the experiments conducted on Database #2. The next subsection presents the experimental results and their analysis obtained by applying this proposed work on Database #3 containing preprocessed images of Database #1.

5.4 Experimental results for Database #3

This section analyzes the experimental results obtained for the experiments conducted on Database #3 containing preprocessed images of Database #1. Tables 8 and 9 list the confusion matrix and performance metrics respectively.

From the Table 8, it can be seen that there is some misclassification in all the classes: knife, handgun and non-weapon classes.

From the Table 9, it can be found that average precision, average recall, average F1 score and average accuracy obtained are 97.57, 97.54, 97.96 and 98.36% respectively. Values of all the performance metrics are greater than 97.50% again confirming the effectiveness of the proposed model.

From the Tables 4, 5, 8 and 9 following conclusions can be drawn:

- Average precision, average recall, average F1 score and average classification accuracy had slightly increased (0.2% approx.) when the proposed preprocessing algorithm is applied on the database confirming the appropriateness of the proposed preprocessing Algorithm 2.
- True positive values for knife class are slightly diminished while true positive values of non-weapon class are improved.

This subsection analyzed the experimental results for Database # 3 which contains preprocessed images of Database #1. Next subsection presents the experimental results and their analysis obtained by applying this proposed work on Database #4.

5.5 Experimental results for Database #4

This section discusses the experimental results obtained for the experiments conducted on Database #4 containing preprocessed images of Database #2.

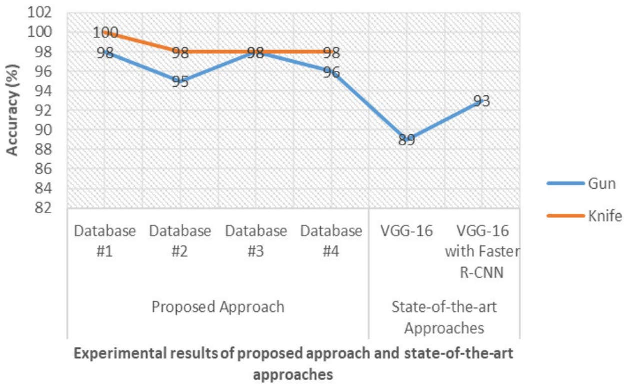

Fig. 7 Accuracy of proposed approach for various databases and state-of-the-art approaches

Table 11 Performance metrics for database #4

	TP	FP	Precision (%)	Recall (%)	F1 score (%)	Accuracy (%)
Knife	401	15	96.39	98.28	97.29	98.20
Handgun	393	07	98.25	96.32	98.23	98.20
Non-weapon	408	00	100.0	100.0	100.0	100.0

Table 12 Accuracy of state-of-the-art approaches and proposed approach

Model	Types of weapons able to identify	Accuracy (%)
VGG-16 [27]	Gun	89.00
VGG-16 with Faster R-CNN [28]	Gun	93.00
Proposed approach	Knife, Gun	98.81

Tables 10 and 11 exhibit the confusion matrix and performance metrics respectively. Confusion matrix shows that all the non-weapon images are correctly classified while there is some misclassification in knife and handgun classes. For the Database #4, average precision, average recall, average F1 score and average accuracy obtained are 98.21, 98.20, 98.50 and 98.80% respectively.

From the Tables 6, 7, 10, and 11 following conclusions are drawn:

- True positive values for knife and handgun classes had increased when the preprocessed data is used.
- Average precision, average recall, average F1 score and average classification accuracy is increased 0.5% approx. when preprocessed data is used.

These observations again confirm the appropriateness and need of preprocessing algorithm for visual weapon identification.

By comparing the experimental results of Database#3 and Database #4, it can be found that the improvement in the results of Database #4 having handheld images is higher as compared to Database #2. It confirms that need of preprocessing algorithm is more for the images having complex background.

The next subsection validates the proposed approach for visual weapon identification by comparing the experimental results of the proposed approach with the state-of-the-art approaches.

5.6 Comparison with existing state-of-the-art approaches

In this section, proposed CNN model is compared with the state-of-the-art approaches for visual weapon identification in terms of average accuracies and ability to identify categories of weapons.

Figure 7 presents the accuracy of proposed approach for all the four databases and for the state-of-the-art approaches. From this figure it can be observed that the performance of proposed approach for both of gun and knife identification is higher than the state-of-the-art approaches. Table 12 summarizes the average accuracy of proposed approach and the state-of-the-art approaches. The average accuracy achieved by proposed approach is on an average 7.8% higher than the others which validate this proposed approach for weapon identification.

This section validated the proposed approach by comparing the experimental outcomes of this proposed approach with the state-of-the-art approaches. The next and final section concludes this paper.

6 Conclusion

This work introduces a framework to identify visual weapons in streaming video surveillance. It is essential for controlling the crime happening in public places. A new CNN model is used for identifying the visual weapons and as well classifying weapon & non-weapon objects in real-time surveillance scenes. This work also presents two new algorithms: one for producing new images similar to input image but having variations in shape, texture, and orientation while other for enhancing the quality of the images. These two algorithms help to solve a large dataset requirement for robust training of the CNN model. Four separate experiments are conducted one each for Database #1, Database #2, Database #3, and Database #4. Here, Database #1 and Database #2 contain isolated and handheld images of weapons and non-weapon class while their preprocessed images are kept in Database #3 and Database #4 respectively. Experimentally, it is found that the overall classification accuracy of the proposed CNN model with Database #1 is 98.07% approx, for Database #2 is 98.42% approx, for Database #3 is 98.36% approx, and for Database #4 is 98.81% approx. Small variations in average accuracy confirm the robustness of this proposed model. The most promising results have been obtained with this proposed CNN model, trained on Database #4, where very less number of false positives, 98.20% recall, 98.21% precision and 98.50% F1 score are obtained. Higher accuracies obtained in case of both categories of images (isolated and handheld) confirm the viability of this proposed approach in intelligent and automated video surveillance. High precision and recall values of the proposed CNN approach confirm the usefulness of this research in being used commercially.

References

1. National Crime Records Bureau. Ministry of Home Affairs. New Delhi, India
2. India third highest in gun-related deaths, firearm mortality rate beats China, Pakistan, Bangladesh: Us study. https://www.counterview.net/2018/09/india-third-highest-in-gun-related.html. Accessed 14 July 2020
3. Velastin, S.A., Boghossian, B.A., Vicencio-Silva, M.A.: A motion-based image processing system for detecting potentially

dangerous situations in underground railway stations. Transport. Res. Part C **14**(2), 96–113 (2006)

4. Ainsworth, T.: Buyer beware. Secur. Oz **19**, 18–26 (2002)
5. Singh, D.K., Kushwaha, D.S.: Ilut based skin colour modelling for human detection. Indian J. Sci. Technol. (2016). https://doi.org/10.17485/ijst/2016/v9i32/92420,
6. Singh, D.K., Kushwaha, D.S.: Tracking movements of humans in a real-time surveillance scene. In: Fifth International Conference on Soft Computing for Problem Solving, pp. 491–500 (2016)
7. Jalal, A., Kamal, S.: Real-time life logging via a depth silhouette-based human activity recognition system for smart home services. In: Eleventh IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), pp. 74–80 (2014)
8. Kumar, T., Kushwaha, D.S.: Traffic surveillance and speed limit violation detection system. J. Intell. Fuzzy Syst. **32**(5), 3761–3773 (2017)
9. Dixit, K.: UP: mela police bans licensed weapons during Kumbh, Times of India, 10 Jan 2019. Accessed 3 Aug 2021
10. Kumar, K.: Text query based summarized event searching interface system using deep learning over cloud. Multimed. Tools. Appl. **80**(7), 11079–11094 (2021)
11. Kumar, K., Shrimankar, D.D.: Deep event learning boost-up approach: DELTA. Multimed. Tools Appl. **77**(20), 26635–26655 (2018)
12. Kumar, K., Shrimankar, D.D.: F-DES: fast and deep event summarization. IEEE Trans. Multimed. **20**(2), 323–34 (2017)
13. Kumar, K.: EVS-DK: event video skimming using deep keyframe. J. Vis. Commun. Image Represent. **58**, 345–352 (2019)
14. Olmos, R., Tabik, S., Herrera, F.: Automatic handgun detection alarm in videos using deep learning. Neurocomputing **275**, 66–72 (2018)
15. Tiwari, R.K., Verma, G.K.: A computer vision based framework for visual gun detection using Harris interest point detector. Procedia Comput. Sci. **54**, 703–712 (2015)
16. Simonyan, K., Zisserman, A.: Very deep convolutional networks for largescale image recognition. arXiv:1409.1556 (2014)
17. Maksimova, A., Matiolanski, A., Wassermann, J.: Fuzzy classification method for knife detection problem. In: International Conference on Multimedia Communications, Services and Security, pp. 159–169 (2014)
18. Glowacz, A., Kmieć, M., Dziech, A.: Visual detection of knives in security applications using Active Appearance Models. Multimed. Tools Appl. **74**(12), 4253–4267 (2015)
19. Cootes, T.F., Edwards, G.J., Taylor, C.J..: Active appearance models. In: European Conference on Computer Vision, pp. 484–498 (1998)
20. Derpanis, K.G.: The Harris corner detector, York University, vol. 2 (2004)
21. Glowacz, A., Kmieć, M., Dziech, A.: Towards robust visual knife detection in images: active appearance models initialised with shape-specific interest points. In: International Conference on Multimedia Communications, Services and Security, pp. 148–158 (2012)
22. Kmiec, M., Glowacz, A.: Object detection in security applications using dominant edge directions. Pattern Recogn. Lett. **52**, 72–79 (2015)
23. Kmiec, M., Glowacz, A.: An approach to robust visual knife detection. Mach. Graph. Vis. **20**(2), 215–227 (2011)
24. Buckchash, H.: Raman, B.: A robust object detector: application to detection of visual knives. In: International Conference on Multimedia and Expo Workshops (ICMEW), pp. 633–638 (2017)
25. Grega, M., Matiolanski, A., Guzik, P, Leszczuk, M.: Automated detection of firearms and knives in a CCTV image. Sensors **16**(1), 47 (2016)
26. Susarla, P., Agrawal, U., Jayagopi, D.B.: Human weapon activity recognition in surveillance videos using structural-rnn. In: Second Mediterranean Conference on Pattern Recognition and Artificial Intelligence, pp. 101–107 (2018)
27. Lai, J., Maples, S.: Developing a real-time gun detection classifier. Stanford University, World Academy of Science, Trieste (2017)
28. Verma, G.K., Dhillon, A.: A handheld gun detection using faster r-cnn deep learning. In: Second International Conference on Computer and Communication Technology, pp. 84–88 (2017)
29. Egiazarov, A. ., Mavroeidis, V., Zennaro, F.M.., Vishi, K.: Firearm detection and segmentation using an ensemble of semantic neural networks. arXiv:2003.00805, 2020
30. Dwivedi, N. , Singh, D.K., Kushwaha, D.S.: Weapon classification using deep convolutional neural network. In: IEEE Conference on Information and Communication Technology, pp. 1–5 (2019)
31. Marquez, E.S., Hare, J.S., Niranjan, M.: Deep cascade learning. IEEE Trans. Neural Netw. Learn. Syst. **29**(11), 5475–5485 (2018)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.