Weapon Classification using Deep Convolutional Neural Network

Neelam Dwivedi
Department of
Computer Science and Engineering
MNNIT Allahabad, Prayagraj
Email: neelamdw@gmail.com

Dushyant Kumar Singh
Department of
Computer Science and Engineering
MNNIT Allahabad, Prayagraj
Email: dushyant@mnnit.ac.in

Dharmender Singh Kushwaha
Department of
Computer Science and Engineering
MNNIT Allahabad, Prayagraj
Email: dsk@mnnit.ac.in

***Abstract*—Increasing crimes in public nowadays pose a serious need of active surveillance systems to overcome such happenings. Type of weapon used in the crime determines its seriousness and nature of crime. An active surveillance with weapon classification can help deciding the course of action while identifying the possibilities of any crime happening. This paper presents a novel approach for weapon classification using Deep Convolutional Neural Networks (DCNN). That is based on the VGGNet architecture. VGGNet is the most recognized CNN architecture which got its place in ImageNet competition 2014, organized for image classification problems. Thus, weights of pre-trained VGG16 model are taken as the initial weights of convolutional layers for the proposed architecture, where three classes: knife, gun and no-weapon are used to train the classifier. To fine tune the weights of the proposed DCNN, it is trained on the images of these classes downloaded from internet and other captured in the lab. Experiments are performed on Nvidia GeForce GTX1050 Ti GPU to achieve faster and exhaustive training on a large image set. A higher accuracy level of 98.41% is achieved for weapon classification.**

***Index Terms*—weapon, deep convolutional neural networks, VGG16, Model A, Model B**

I. Introduction

Nowadays, many cases of crimes are reported in public places / homes using different types of weapons such as firearms, swords, cutters, etc. To monitor and minimize such types of crimes, CCTV camera is installed in public places. Generally, the video footages recorded through these cameras are monitored by security staff. Success and failure of detecting crime depends on the attention of operator. It is not always possible for a person to pay attention on all the video feeds on a single screen recorded through multiple video cameras. According to Velastin et al. [1], a CCTV operator is not able to actively recognize the objects present in video feeds after 20 to 40 minutes. According to another study published in Security Oz Magazine [2], Miss rate of an object by the operator increases up to 95% after 22 minutes. Thus, to overcome these limitations there is a need of automatic video surveillance systems which can detect weapons before the start of any inhuman activity. Besides this, in most of the crimes one or more weapons are used in the attacks. So, weapon detection capabilities of an automated surveillance system increase the level of security. A robust weapon classification system can greatly enhance the efficacy of video surveillance [3][4], smart homes [5], intrusion detection [6], detection of security breaches in smart cities [7][8] and other application domains. Nature and extent of crime depends on the types of weapon that is used. If an automated video surveillance system has the ability to generate a prior alert then by timely reaction losses may be reduced to the maximum extent. Advantage of weapon classification can also be added to an automated surveillance system. Talking on the weapon type it has been seen that guns and/or knives are majorly used in the crime because they are ease to carry. Therefore, the work proposed here concentrates on classifying or detecting these two classes of weapons in any image.

Weapons may be classified either using standard techniques with machine learning classifier or by using deep learning based techniques. In the standard techniques, features are extracted manually and are used to train the classifier (model). The trained model is further used for classifying any new input image. Accuracy of such types of approaches depends on the robustness and diversity of extracted features. To overcome these limitations, deep Convolutional Neural Networks is better to be used as it does not require any explicit feature of the input image.

Deep Convolutional Neural Networks consist of a number of convolutional layers, pooling layers and fully connected layers. Convolutional layers extract various features from the input image which includes a high degree of invariance, scaling and other forms of deformation. Thereafter, fully connected layers learn from these features. Due to this property, deep CNN is applied in many applications and achieves better accuracy in comparison to standard machine learning based approach. One more benefit of using such architecture is that it can be partially or fully reused for related applications. It can be done by using the concept of transfer learning which reduces model development time and also the requirement of large dataset. Due to these advantages of deep CNNs architecture, we have initialized the weights of convolutional layers of new model with the weights of pre-trained VGG16 model for the weapon classification. VGG16 architecture was the first runner-up in ILSVR (ImageNet Large Scale Visual Recognition Competition) 2014 [9]. For this, the architecture was trained with Imagenet dataset for 1000 classes of images.

The remaining part of the paper is organized as follows: In

978-1-7281-5398-8/19/$31.00

Section II, related work is presented. Proposed approach for weapon classification is explained in Section III. Experimental results are discussed in Section IV. Section V concludes the paper followed by references.

II. Related work

In this section, a brief review of state-of-the-art approaches for weapon detection, classification and deep neural networks is presented. Many researchers are working in the area of weapon detection and classification. Maksimova et al. [10] presented a classification model for knife detection. In this model, fuzzy clustering approach is used to detect the knife present in any frame of a video. Tiwari and Verma [11] proposed a visual gun detection frame work for automatic surveillance. Color based segmentation and harris interest point detector is used to detect the gun in images. Their proposed method takes long processing time which is not suitable for real time application. Performance of this approach is also not good with real time videos having changes in the illumination. Buckchash and Raman [12] proposed an object detection algorithm to detect and classify a knife. Their proposed algorithm works on three phases: Mixture of Gaussians for foreground object detection, key point detectors for object localization and Multi-Resolution Analysis for object classification. Grega et al. [13] proposed an algorithm to detect a firearm or knife in the image and to generate some alert for human. They also tried to reduce the number of false alarms for real-life surveillance videos. Glowacz et al. [14] have presented a knife detection method in the images. This method is based on active appearance models and harris corner detector. Harris corner detector is used to find the tips of knife. Thereafter, active appearance model of knife is created using these tips. The overall performance of this architecture depends on harris corner detector accuracy.

Susarla et al. [15] presented a Structural Recurrent Neural Networks (SRNN) model to recognize the spatio-temporal human-object interactions in video surveillance. Olmos et al. [16] presented a pistol detection system in a video based on faster R-CNN. The authors created their own data-set for the training and testing purpose. Further, they reformulate their detection problem to minimize false positive rate. Ghazi et al. [17] applied transfer learning method to identify plant species in an image using deep convolutional neural networks. Here, plant task dataset is used to fine-tune the pre-trained deep learning models. Furthermore, to improve the performance different classifiers are fused together and also adjust the different parameters of the model.

Existing approaches in the literature for weapon classification having low accuracy and also detect one category of weapon only. To improve the accuracy, new architecture is proposed in this paper and also analyze the effect of changes in the dropout rate and number of neurons in the network. In the next section, proposed architecture for weapon classification is discussed.

TABLE I
Architectures of VGG-16, Model A and Model B

Input size	VGG 16	Model A	Model B
160	3 × 3, 64	3 × 3, 64	3 × 3, 64
	3 × 3, 64	3 × 3, 64	3 × 3, 64
	2 × 2, pool	2 × 2, pool	2 × 2, pool
80	3 × 3, 128	3 × 3, 128	3 × 3, 128
	3 × 3, 128	3 × 3, 128	3 × 3, 128
	2 × 2, pool	2 × 2, pool	2 × 2, pool
40	3 × 3, 256	3 × 3, 256	3 × 3, 256
	3 × 3, 256	3 × 3, 256	3 × 3, 256
	3 × 3, 256	3 × 3, 256	3 × 3, 256
	2 × 2, pool	2 × 2, pool	2 × 2, pool
20	3 × 3, 512	3 × 3, 512	3 × 3, 512
	3 × 3, 512	3 × 3, 512	3 × 3, 512
	3 × 3, 512	3 × 3, 512	3 × 3, 512
	2 × 2, pool	2 × 2, pool	2 × 2, pool
10	3 × 3, 512	3 × 3, 512	3 × 3, 512
	3 × 3, 512	3 × 3, 512	3 × 3, 512
	3 × 3, 512	3 × 3, 512	3 × 3, 512
	2 × 2, pool	2 × 2, pool	2 × 2, pool
FC1	4096	1024	2048
FC2	4096	512	1024
FC3	1000	3	3

III. Proposed Approach for Weapon Classification

This paper presents deep CNNs architecture for weapon classification based on VGGNet. Proposed approach uses the concept of transfer learning for initializing the weights of convolutional layer of the proposed architecture. Transfer learning is a method of learning where knowledge obtained from one scenario can be used to improve the learning of any another related scenario. Two different models have been proposed by changing the number of neurons in the fully connected layer to examine the effect of number of neurons on the classification accuracy. Convolutional layers of both the models (Model A and Model B) are same as VGG16 model. Model A and Model B both contain three fully connected layers named as FC1, FC2 and FC3. Model A contains 1024 neurons, 512 neurons and 3 neurons in FC1, FC2 and FC3, respectively. Model B contains 2048 neurons, 1024 neurons and 3 neurons in FC1, FC2, and FC3 layers respectively. The neurons of first two fully connected layers (FC1 and FC2) are used to train the model. The number of neurons in last layer (FC3) is used to classify the objects. Since all the objects are to be classified among three classes so the number of neurons used for FC3 in both the models is 3. Table I shows the detail architecture of VGG16, Model A and Model B.

Both models consist of five groups of convolutional layer followed by three fully connected layers. First two groups contain two convolutional layers each while last three groups consist of three convolutional layers each. These thirteen convolutional layers are used for extracting the features of the input images. Later, all fully connected layers learn from these extracted features. Thereafter, Softmax activation function is applied to get the probabilistic distribution of the predicted class. Weights of a convolutional layer of the proposed architecture are initialized with the weights of convolutional layers of the pre-trained VGG16 model. Weights of fully

connected layers are initialized randomly. For fine tuning of these weights, network is trained with the selected samples of different classes of images belonging to knife class, gun class and no-weapon class. Experimental setup and results have been explained in the section IV.

Experimental setup and results have been explained in the section IV.

IV. Experiments and Their Analysis

Experiments have been conducted for three classes of images: knife, gun and no-weapon class. Various types of knives have been downloaded from the internet and some of the pictures were captured from the camera in our lab. Gun class consists of different types of pistol images downloaded from the internet. No-weapon class contain the images of human beings, cars, chairs etc. Dataset is divided randomly into two sets: training and test. Training set contains 1520, 1800 and 1176 images of knives, guns and no-weapons class respectively. Test set contains 344, 368 and 296 images of knives, guns and no-weapons classes respectively. Figure 1 shows sample input images of knife, gun and no-weapon class.

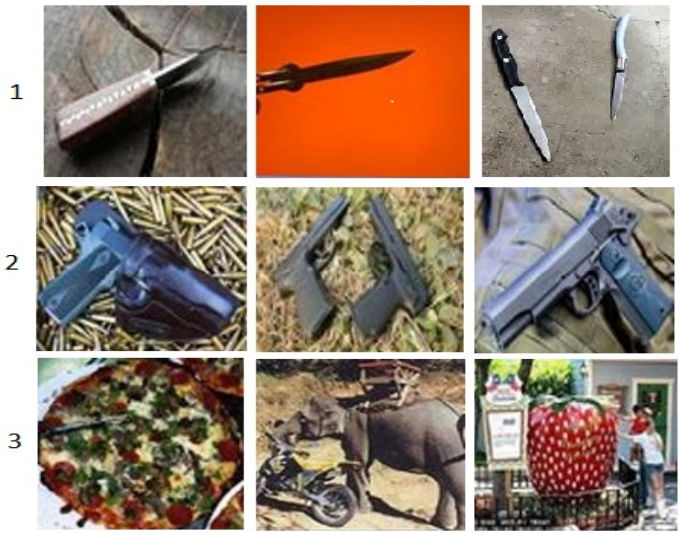

Fig. 1. Input images: (1) shows samples of knives,(2) shows samples of guns and (3) shows samples of no-weapon class.

Since dataset is small, so for proper training of the model following changes have been made:

- Fully connected layer is shrunk sharply to reduce the number of parameters.
- Dropout value is increased for preventing the network to be over fitted.

To check the effects of dropout rate two separate experiments have been conducted for both the models namely Model A and Model B. For all of the experiments, '20' epochs have been used for the training of these networks. Batch size of '8' is used for these experiments in the training phase. After the processing of 8 images weights of convolutional layer of our model are updated. Here, input size of an image is taken as 160 $\times$ 160. Two consecutive convolutional operations are performed on the input image by applying 64 kernels each of size 3 $\times$ 3, to extract the initial features of the image. A pooling layer of size 2$\times$ 2 is applied later to reduce the size of an input image. Now, the input image size is reduced from 160 $\times$ 160 to 80 $\times$ 80. Two consecutive convolutional operations are again performed on the obtained image by applying 128 kernels followed by applying a max pool operation to get some deeper feature. To obtain deeper feature of the image, again three consecutive convolutional operations are performed by applying 256 kernels followed by max pool operation. This is repeated two more times by using 512 kernels consecutively. These features are used to train the fully connected layers: FC1, FC2 and FC3 for both the models A and B. To update the weights of the network, Adaptive Moment Estimation (Adam) algorithm is used. To check the effect of dropout on the network, both models have been trained using two different dropout values: 0.5 and 0.7.

Confusion matrix, classification accuracy and precision-recall curve are used as the objective evaluation parameter to analyze the efficiency of the proposed architectures. Figure 2 shows sample output of the proposed deep CNN architecture.

Fig. 2. Output images: (1) shows samples of knives, (2) shows samples of guns and (3) shows samples of no-weapon class.

TABLE II
Confusion matrix for Model A with dropout rate=0.5

	Knife	Gun	No weapon
Knife	342	0	2
Gun	0	366	2
No weapon	7	5	284

TABLE III
CONFUSION MATRIX FOR MODEL B WITH DROPOUT RATE=0.5

	Knife	Gun	No weapon
Knife	337	4	3
Gun	2	365	1
No weapon	6	5	285

Table II shows the confusion matrix for Model A when dropout rate is taken as 0.5 while Table III shows the confusion matrix for Model B with the same dropout rate. Model B has 2048 and 1024 neurons in the FC1 and FC2, respectively which is double from the neurons (1024 and 512) of Model A. From Table II and Table III, it can be seen that in Model B, the number of true positives increases only for no-weapon class and decreases for the two remaining classes. Thus, it can be concluded from these results that classification accuracy not necessarily increases when number of neuron is increased.

TABLE IV
CONFUSION MATRIX FOR MODEL A WITH DROPOUT RATE=0.7

	Knife	Gun	No weapon
Knife	342	0	2
Gun	2	365	1
No weapon	16	7	273

TABLE V
CONFUSION MATRIX FOR MODEL B WITH DROPOUT RATE=0.7

	Knife	Gun	No weapon
Knife	341	1	2
Gun	1	366	1
No weapon	9	16	271

Table IV shows the confusion matrix for Model A when the dropout rate is taken as 0.7 while Table V shows the confusion matrix for Model B with the same dropout rate. From the Table IV and Table V, it can be seen that only for Gun class True Positive value is increased while for other classes it is decreased for Model B.

From the Table II and Table IV, it can be seen that when dropout rate is increased for Model A then number of true positive decreases. From the Table III and Table V, it can be seen that when the dropout rate is increased, then number of true positive increases in Model B for knife and gun classes while decreased for no-weapon class. From these two observations, it can be concluded that increase or decrease in True Positive of any class is solely not dependent on increase or decrease in the dropout rate.

From Table II, Table III, Table IV and Table V following conclusions have been made.

- By increasing the number of neurons in the fully connected layers, it is not necessary that the total number of true positives will also increase for all the classes.
- Number of true positives may or may not decrease with increasing the dropout rate.
- Further extended relation between different model parameters and evaluation measures are the part of forthcoming experiments.

TABLE VI
CLASSIFICATION ACCURACY

	Accuracy (%)	
	Dropout = 0.5	Dropout =0.7
Model A	98.41	97.22
Model B	97.91	97.02

Table VI shows the accuracy of both the Models: Model A and Model B, for two dropout rates 0.5 and 0.7. It can be concluded from this table that average accuracy decreases with increased dropout rate.

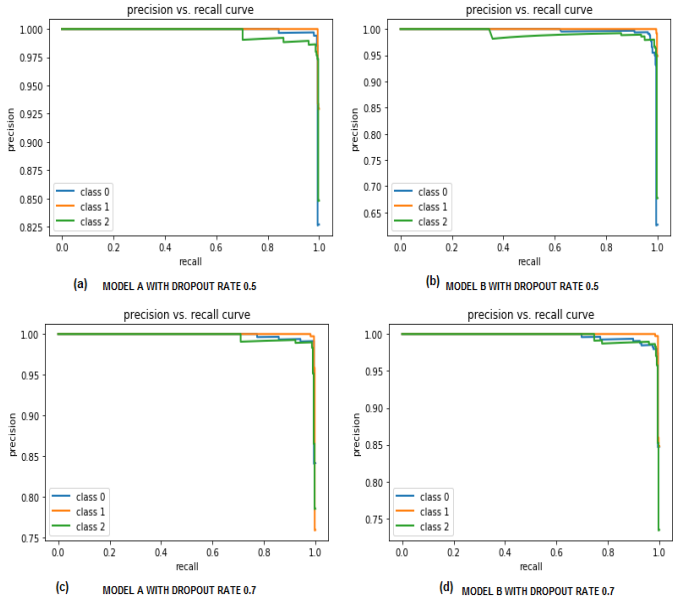

Fig. 3. (a) PR curve for Model A with dropout rate = 0.5 (b) PR curve for Model B with dropout rate = 0.5 (c) PR curve for Model A with dropout rate = 0.7 (d) PR curve for Model B with dropout rate = 0.7

Figure 3(a) and 3(b) show the Precision-Recall (PR) curve with dropout rate = 0.5 for Model A and Model B, respectively while, Figure 3(c) and 3(d) show the respective PR curve for Model A and Model B with dropout rate = 0.7. These PR curves show the retreival accuracy of respective models for all the classes. Larger values of Area Under Curve (AUC) depicts better retrieval accuracy. From these PR curves it can be seen that AUC is $\approx$ 1 for all of these models for all the classes. From the Table VI and Figure 3, it can be concluded that the classification accuracy as well as retreival accuracy are close to 99% for the proposed models confirming the appropriateness of this work.

V. CONCLUSION

Weapon detection and classification with the help of video surveillance is required to reduce the crimes that happens in public places like schools, malls etc. This paper proposes

two new deep CNN architectures by considering VGG16 as a base model. To train the proposed networks, weights of convolutional layer is initialized with the weights of pre-trained VGG16 model on Imagenet dataset while weights of fully connected layer are randomly initialized. Weights of the proposed network are fine tuned by training this network with the images of knives, guns and no-weapons classes. Maximum accuracy of 98.41% is obtained for Model A with 0.5 dropout rate proving the appropriateness of the proposed approach. Model A with dropout of 0.5 has less number of neurons as compared to Model B that has 0.7 dropout in fully connected layers. Through this result, it is concluded that accuracy will always not be improved just by increasing the number of neurons. Experimentally, it is also concluded that average accuracy decreases with increased dropout rate.

REFERENCES

[1] Sergio A Velastin, Boghos A Boghossian, and Maria Alicia Vicencio-Silva. A motion-based image processing system for detecting potentially dangerous situations in underground railway stations. *Transportation Research Part C: Emerging Technologies*, 14(2):96–113, 2006.

[2] Trevor Ainsworth. Buyer beware. *Security Oz*, 19:18–26, 2002.

[3] Dushyant Kumar Singh and Dharmender Singh Kushwaha. Ilut based skin colour modelling for human detection. *Indian J. Sci. Technol*, 9:32, 2016.

[4] Dushyant Kumar Singh and Dharmender Singh Kushwaha. Tracking movements of humans in a real-time surveillance scene. In *Proceedings of Fifth International Conference on Soft Computing for Problem Solving*, pages 491–500. Springer, 2016.

[5] Ahmad Jalal and Shaharyar Kamal. Real-time life logging via a depth silhouette-based human activity recognition system for smart home services. In *2014 11th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pages 74–80. IEEE, 2014.

[6] Dushyant Kumar Singh and Dharmender Singh Kushwaha. Automatic intruder combat system: a way to smart border surveillance. *Defence Science Journal*, 67(1):50–58, 2017.

[7] Tarun Kumar and Dharmender Singh Kushwaha. Traffic surveillance and speed limit violation detection system. *Journal of Intelligent & Fuzzy Systems*, 32(5):3761–3773, 2017.

[8] Tarun Kumar and Dharmender Singh Kushwaha. An efficient approach for detection and speed estimation of moving vehicles. *Procedia Computer Science*, 89:726–731, 2016.

[9] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[10] Aleksandra Maksimova, Andrzej Matiolański, and Jakob Wassermann. Fuzzy classification method for knife detection problem. In *International Conference on Multimedia Communications, Services and Security*, pages 159–169. Springer, 2014.

[11] Rohit Kumar Tiwari and Gyanendra K Verma. A computer vision based framework for visual gun detection using harris interest point detector. *Procedia Computer Science*, 54:703–712, 2015.

[12] Himanshu Buckchash and Balasubramanian Raman. A robust object detector: application to detection of visual knives. In *2017 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, pages 633–638. IEEE, 2017.

[13] Alberto Castillo, Siham Tabik, Francisco Pérez, Roberto Olmos, and Francisco Herrera. Brightness guided preprocessing for automatic cold steel weapon detection in surveillance videos with deep learning. *Neurocomputing*, 330:151–161, 2019.

[14] Michał Grega, Andrzej Matiolański, Piotr Guzik, and Mikołaj Leszczuk. Automated detection of firearms and knives in a cctv image. *Sensors*, 16(1):47, 2016.

[15] Praneeth Susarla, Utkarsh Agrawal, and Dinesh Babu Jayagopi. Human weapon-activity recognition in surveillance videos using structural-rnn. In *Proceedings of the 2nd Mediterranean Conference on Pattern Recognition and Artificial Intelligence*, pages 101–107. ACM, 2018.

[16] Andrzej Glowacz, Marcin Kmieć, and Andrzej Dziech. Visual detection of knives in security applications using active appearance models. *Multimedia Tools and Applications*, 74(12):4253–4267, 2015.

[17] Roberto Olmos, Siham Tabik, and Francisco Herrera. Automatic handgun detection alarm in videos using deep learning. *Neurocomputing*, 275:66–72, 2018.