International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-8 Issue-4, April 2019

Computer Vision Framework for Visual Sharp Object Detection using Deep Learning Model

Nitesh Ramakrishnan, Anandhanarayanan Kamalakannan, Balika J Chelliah, Govindaraj Rajamanickam

Abstract - Deep learning models are widely used for visual image feature extraction and classification. Troublemakers in human society may handle sharp objects like knifes, blades to perform crimes like burglary in public places. To monitor such activities, visual sharp object detection software needs to be integrated with camera based security and surveillance systems. To implement this application, our paper discusses about computer vision framework for sharp object detection using CNN model. Initially, object detection model was built using different CNN architectures namely AlexNet, ZFNet and VGG13. In order to improve the training and testing accuracy of the above models, a new CNN model was proposed with modified VGG architecture. The proposed CNN model has limited number of convolution layers with minimum weight parameters. Thus this model improves computation efficiency when executed on Intel CPUs and delivers better accuracy in training and testing when compared with other CNN architectures. Around 98% training and 92.2% testing accuracy was obtained for this model.

Keywords - Convolutional Neural Network (CNN); Central Processing Unit (CPU); Graphical User Interface (GUI); Sharp Object Detection; Image Data Preparation

I. INTRODUCTION

Image classification is considered as main task in pattern recognition problem. For security purposes in public places like election booth, airports and railway stations etc., a visual object detection software tool needs to be developed for identifying sharp objects like knives and blades. Development of computer vision framework for sharp object detection helps the security officials monitor the happening of major crimes in crowded places. The sharp object detection algorithm can be implemented using different pattern description techniques like edge detection, line detection and boundary representation [1] etc. Geometric features extracted from existing pattern description methods like canny, sobel, prewitt, hough transform and chain codes do not provide precise features for object recognition.

Manuscript published on 30 April 2019.
* Correspondence Author (s)
Nitesh Ramakrishnan, Student Trainee, Central Electronics Engineering Research Institute, Chennai Centre, CSIR Campus, Chennai-113, India.
Anandhanarayanan Kamalakannan, Central Electronics Engineering Research Institute, Chennai Centre, CSIR Campus, Chennai-113, India.
Balika J Chelliah, Dept. of Computer Science & Engineering, SRM Institute of Science and Technology, Chennai-89, India.
Govindaraj Rajamanickam, Central Electronics Engineering Research Institute, Chennai Centre, CSIR Campus, Chennai-113, India

In order to improve object recognition, deep learning based pattern recognition techniques can be used because they do not need separate feature extraction methods as mentioned above [1]. The paper [2] discusses about implementation of a computer vision framework for visual gun detection using harris corner detector. Our paper aims to build a visual sharp object detection framework using CNN model. Mostly deep learning models are highly suitable for computer vision applications [3] due to its excellent object recognition accuracy found during training and testing. In order to develop the detection algorithm, different lightweight CNN models based on AlexNet [4], ZFNet and VGG13 architectures were trained and tested on CPU. The AlexNet and ZFNet models delivered around 85% training and 80% testing accuracy. The VGG13 architecture delivered 93% training accuracy but there was no major improvement in the testing accuracy. To enhance the detection performance of the above architectures further; a new CNN model based on the existing VGG architecture was employed with 8 convolution and 3 fully connected layers having only 62.6 million training parameters. This model has a limited number of convolution layers and possesses less number of weight parameters when compared with VGG16 and VGG13 architectures. The training and testing accuracy of the proposed model was found to be 98% and 92.2% respectively which is better than the results of other architectures as discussed above.

The paper is arranged as follows: Section II discusses about the challenges incurred by running larger CNN architectures like VGG16, ResNet, and GoogLeNet on computing system. Section III describes about the image dataset preparation and the proposed CNN architecture used in the sharp object detection framework. The CNN model implementation and computer vision GUI for the proposed framework is explained in section IV. In section V, the training and testing results of different CNN models are discussed. The conclusion is given in section VI.

II. CHALLENGES

Carrying sharp objects like knives, scissors and blades are mostly prohibited in vital places. Manual detection of sharp objects in CCTV screen using the naked eye is not possible all the time due to human fatigue. With the help of the proposed framework, an alarm can be raised when a potential threat [5] is found on the monitor. Captured images of hand held sharp objects show limited variation in their appearance and shape.

Published By:
Blue Eyes Intelligence Engineering & Sciences Publication (BEIESP)

Therefore it is enough to implement sharp object detection and classification algorithm efficiently with limited dataset.

Recent CNN based image classification techniques use VGG16 [6], ResNet, Faster R-CNN [7] and Inception V3 models [8], which require high-performance CUDA enabled GPU cards and larger image dataset. Although these methods have given results, it is not optimal for our task, since its processing time and computation resource requirements are high. Our proposed CNN model can run efficiently in a CPU system with minimum convolution layers and limited parameters without any additional GPU card.

III. METHODOLOGY

CNN based pattern recognition model has made a huge breakthrough in the field of image classification and detection [9]. It is widely used in image processing and computer vision because it can extract high-level image features and stores it for future analysis. These models are trained with datasets which contain labeled images of different classes. During the training process, the CNN is trained to recognize and learn the image patterns belonging to different classes. The proposed CNN model is a modified version of VGG architecture with 8 convolution layers and 3 fully connected layers. Due to its lightweight architecture, it can be trained and executed smoothly on a CPU machine. This model implementation also saves us from buying costlier high-performance workstation PC supported by GPU cards. This model was chosen due to three main advantages namely, good detection accuracy, less training parameters and runs on minimum computation hardware. It consists of an input layer, multiple convolutional layers, max-pooling layers, fully-connected layers followed by an output layer with 3 nodes to classify different objects like knife, utensil and water bottles made of metal components.

A. Image Dataset Preparation

Image dataset is a collection of labeled sample images belonging to different classes which are considered for training and testing purpose. The dataset plays a vital role in improving the accuracy of deep learning techniques. To obtain better training accuracy, the dataset must be refined before it is processed by the CNN model. This is because proper training of the CNN relies purely on the image dataset quality given to it. The sample knife image dataset used for training purpose is shown in fig.1.

Fig.1 Sample images from knife dataset (Courtesy:

In this experiment, around 300 images of knife along with other objects were taken from internet image sources [10]. The image dataset consists of three classes namely knife, utensil and water bottle. The sample size of the image dataset was increased from 300 to 1500 by performing data augmentation techniques like horizontal flipping, vertical flipping, translation, rotation and rescaling etc. Thus this technique satisfies the input data requirement for training the CNN model. Around 1000 augmented images were used for training and the remaining 500 were used for testing. When we generate image dataset, all the images will not be of uniform size, brightness, orientation, etc. These are the factors which reduce the number of images chosen for training dataset. Uniformity in training images is very much important for better model prediction. Usually CNN model will take training images with specified input size, so we need to resize all the training images to maintain uniformity according to the model input specification. This preprocessing step is very important in any deep learning technique because it enhances the model accuracy and testing performance.

B. Proposed CNN architecture

The proposed CNN model was implemented to gain maximum accuracy with the limited dataset. This algorithm was developed by modifying the current VGG16 architecture by reducing the number of convolution layers [11] in each block and also the number of convolution filters applied in each layer is reduced to half. The reduction in convolution filters drastically reduces the number of weight parameters needed to train the CNN model. Therefore this lightweight CNN model can be executed on CPU.

Convolution layer in CNN model is constructed with a varying number of convolution filters with fixed kernel size. The convolution filter is traversed across the width and height of the image such that each convolution layer in the CNN model extracts low level and high-level features, including lines and shapes automatically. In each block, the max pooling layer followed by convolution layer gradually reduces or halves the input size feed to the next block as shown in the proposed CNN model fig.2.

During CNN model training, augmented image dataset is passed as input through a series of blocks consisting of convolution and max-pooling layers. The input size of the proposed CNN model is fixed at 224x224. A 3x3 kernel size is used in all the convolutional layers to extract 2D image features. In this CNN architecture, 5 blocks are defined with varying number of convolution and max-pooling layers. In the 1st and 2nd block, single convolution layer was defined with 32 filters and 64 filters respectively. In the 3rd, 4th, and 5th blocks two convolution layers are defined with filter size 128, 256 and 256 respectively. In each block, single max pooling layer with 2x2 kernel size follows the preceding convolution layers and the extracted features are quantized to half size. At the end of the 5th block, a 7x7 image feature is obtained which is then flattened and passed to three fully connected layers. Third fully connected layer possess 1000 output nodes which are connected to three softmax output. The number of softmax nodes defines the number of classes used for classification and detection. The proposed CNN model nearly takes 62.6 million parameters for training and testing. Overall this model could run smoothly on new Intel CPU and delivers good accuracy results during training and testing as discussed in Section V.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication (BEIESP)

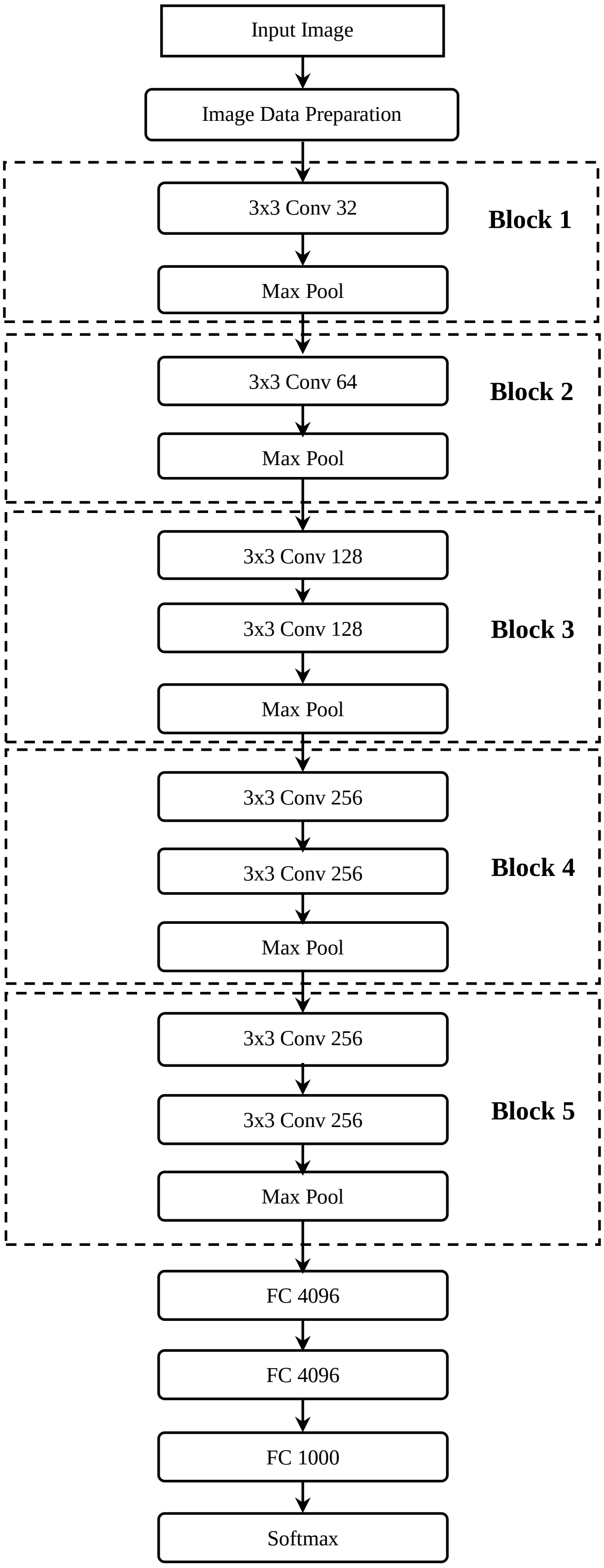


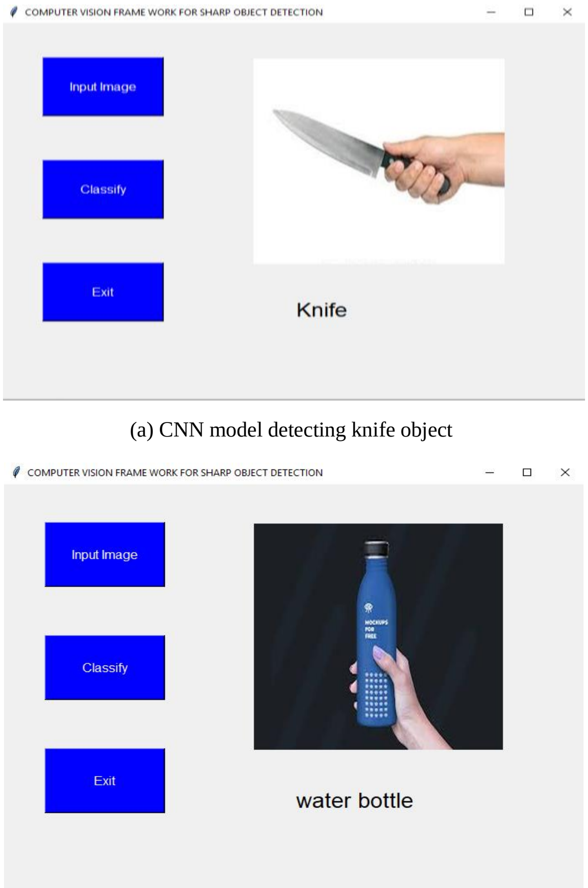
Fig. 2 Proposed CNN

IV. IMPLEMENTATION OF COMPUTER VISION FRAMEWORK

The proposed computer vision framework aims to detect and classify any sharp objects present in a visual image. This framework was implemented using TensorFlow in Python's Integrated Development and Learning Environment (IDLE). To implement the sharp object detection algorithm, the CNN model is trained in CPU by reducing the number of training parameters to minimum, which makes it optimal for most CPU users for real-time implementation. In order to train and test this CNN model, Intel Core i7-4770 CPU 3.4GHz, 4GB RAM, and Microsoft Windows 10 64-bit operating system was used.

A. GUI

GUI framework was designed using python Tkinter library as it is user-friendly and simple. The GUI window consists of three buttons, one for selecting an input image, another one for classifying and detecting objects, the last button was used to close the application. The selected input image is displayed in the main window as shown in fig.3. The sharp object detection model is executed when classify button is clicked. The detection results are displayed in label format as shown in fig.3a & fig.3b.



(a) CNN model detecting knife object

(b) CNN model detecting water bottle

Fig. 3 GUI of computer vision framework for sharp

B. Algorithm Implementation

The object detection algorithm was fundamentally developed in python language using Python's IDLE. To define various layers in the proposed CNN algorithm, open-source machine learning libraries namely TensorFlow and Keras were used. From the generated image dataset, the images were grouped under two main categories: training and testing. The image samples in the dataset belong to three main classes: knife, utensil and water bottle.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication (BEIESP)

The training dataset consists of 1000 images distributed among three classes. Similarly testing dataset consists of 500 images distributed among three classes. The CNN model was trained and validated using training dataset for several epochs. The trained model weights were stored in HDF5 format. Computer vision framework loads the saved CNN model/classifier and detects sharp objects in the newly generated images. The predicted results are displayed in a label box to the end-user. The training and testing accuracy of different CNN models were evaluated referring to the accuracy equation [2][8].

V. RESULTS AND DISCUSSION

The training and testing results of different CNN architectures such as AlexNet, ZFNet, and VGG13 applied over visual sharp object image dataset were studied. AlexNet and ZFNet models have a training accuracy around 85% and testing accuracy around 80%. In order to improve the results, VGG13 layer architecture was trained and tested. The VGG13 layer model achieved 93% training accuracy, but testing accuracy did not show any major improvement with respect to the previous models. To gain better performance in accuracy, a lightweight 11 layer CNN model based on VGG architecture was proposed. The model was tested on 500 images. It is able to detect 241 knives out of 250, 99 utensils out of 125 and 121 water bottles out of 125 images. The proposed model achieved 98% training accuracy and 92.2% overall testing accuracy. For sharp object detection, the performance of the proposed model was found to be better when compared with the other models as shown in Table-1.

Table I. Training and Testing accuracy of

CNN Architecture	Training	Testing
AlexNet	85 %	75.6 %
ZFNet	87 %	80.0 %
VGG13 layers	93 %	82.2 %
Proposed CNN Model (11 layers)	**98 %**	**92.2 %**

VI. CONCLUSION

In this paper, an object recognition algorithm has been proposed to classify and detect objects such as sharp edged knife, utensil, and water bottle present in visual images using different CNN models namely AlexNet, ZFNet and VGG13. In order to improve the training and testing accuracy of the existing models, a new CNN model was proposed with modified VGG architecture. The proposed model executes faster in Intel CPU because of limited weight parameters. The training and testing accuracy results of different models were analyzed on Intel CPU and the results were tabulated. The proposed CNN model gave better performance in terms of training and testing accuracy when compared with other models. This model can also be trained to detect different sharp objects present in visual images.

ACKNOWLEDGEMENT

The research activity has been carried out under student trainee program of Central Electronics Engineering Research Institute, Chennai Centre, CSIR Madras Complex, Taramani, Chennai, India. The author N.R sincerely thanks the Director, CEERI, Pilani and Scientist-In-Charge, CEERI Chennai Centre for extending all the lab facilities to carry out the research work. He also wishes to thank the Head of Department (Computer Science & Engineering), SRM Institute of Science and Technology, Ramapuram, Chennai for granting permission to pursue research work at the CEERI Chennai Centre.

REFERENCES

1. Rafael C. Gonzelez, Richard E. Woods, “Chapter 10 - Image Segmentation”, Digital Image Processing (Second Edition), Prentice Hall of India, pp. 568-572, 2015.
2. Rohit Kumar Tiwari, Gyanendra K. Verma, “A Computer Vision based Framework for Visual Gun Detection using Harris Interest Point Detector,” Procedia Computer Science, vol. 54, pp. 703–712, 2015.
3. Feng Zhou, Weiwei Fan, Qiangqiang Sheng, and Mingliang Tao, “Ship Detection Based On Deep Convolutional Neural Networks For Polsar Images”, IGARSS, pp. 681-684, 2018.
4. Jing Sun, Xibiao Cai, Fuming Sun and Jianguo Zhang, ”Scene Image Classification method based on AlexNet Model”, 3rd International Conference on Informative and Cybernetics for Computer Social Systems, 2016.
5. Nadia Jmour, Sehla Zayen, and Afef Abdelkrim,” Convolutional Neural Networks for image classification”, IEEE ICCABS, 2018.
6. Justin Lai, Sydney Maples, “Developing a Real-Time Gun Detection Classifier”, Stanford University Reports:CS231n, 2017.
7. Hai Duong Nguyen, Young Chul Kim, and Soo Hyung Kim, ” A Method for Fingertips Detection using RGB-D Image and Convolution Neural Network”, 13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery, 2017.
8. Narit Hnoohom and Sumeth Yuenyong, “Thai Fast Food Image Classification Using Deep Learning”, 15th International Conference Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology, 2018.
9. Krizhevsky, A., Sutskever, L., and Hinton, G., “Imagenet classification with deep convolutional neural networks,” Advances in neural information processing systems, 2012.
10. https://www.google.com/search?q=knife+images&newwindow=1&source=lnms&tbm=isch&sa=X&ved=0ahUKEwjajPq7s8LhAhXGVbwKHVM8B6YQ_AUIDigB&biw=1366&bih=657
11. Karen Simonyan, Andrew Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition”, ICLR, 2015, http://arxiv.org/abs/1409.1556.

AUTHORS PROFILE

Nitesh Ramakrishnan, is a final year B.Tech student trainee of CSIR-CEERI, Chennai Centre. He is pursuing Bachelor degree in Computer Science & Engineering from SRM Institute of Science and Technology, Ramapuram, Chennai. He has keen interest in machine learning, database management, image processing and cryptography.

Anandhanarayanan Kamalakannan, is working in CSIR-Central Electronics Engineering Research Institute, Chennai Centre as Technical Officer. He has been actively engaged in executing projects pertaining to X-ray image processing for industrial applications. He received B.E degree in Electronics & Communication Engineering from University of Madras, Chennai in 2004. He then received Master degree in Engineering (Medical Electronics) from College of Engineering, Anna University, Chennai in 2018. His areas of interest are X-ray imaging, artificial intelligence, parallel computing and machine vision.

Balika J Chelliah, is an Associate Professor in Department of Computer Science & Engineering, SRM Institute of Science and Technology, Ramapuram, Chennai. He received his Master and Ph.D degrees in Computer Science & Engineering from SRM Institute of Science and Technology, Chennai. His areas of interest are Service oriented Architecture, Web services, Cloud services and Software Engineering.

Govindaraj Rajamanickam, is working in CSIR-Central Electronics Engineering Research Institute (CEERI), Chennai Centre as Chief Scientist, has more than two decades of R&D experience and engaged in the design and development of electro-optical sensors and imaging technology pertaining to industrial applications. He received M.Sc. degree in Physics from St.Joseph's College (Autonomous), affiliated to Bharathidasan University, Trichy, Tamil Nadu, India in the year 1984. He then received Ph.D. degree in Physics from University of Madras, Chennai in the year 1996. His research interests include optical sensors, imaging spectroscopy, image processing and machine vision systems.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication (BEIESP)