IJCSE International Journal of Computer Sciences and Engineering Open Access

Research Paper Vol.-7, Issue-5, May 2019 E-ISSN: 2347-2693

Real Time Object Identification Using Neural Network with Caffe Model

Anjali Nema[1*], Anshul Khurana[2]

[1,2]CSE, SRIT, RGPV, Jabalpur, India

[*]Corresponding Author: anjali.nnema @gmaill.com, Tel.: +91-8839957983.

DOI: https://doi.org/10.26438/ijcse/v7i5.175182 | Available online at: www.ijcseonline.org

Accepted: 13/May/2019, Published: 31/May/2019

Abstract— Neural Networks has become one of the most demanded areas of Information Technology and it has been successfully applied to solving many issues of Artificial Intelligence, for example, speech recognition, computer vision, natural language processing, and data visualization. This thesis describes the developing the neural network model for object detection and tracking. With the progress of science and technology, information technology was advancing rapidly. The understanding of moving object based on vision has also developed rapidly. Its related technologies have been widely used in public transportation, square, government, bank and other scenes. At present, there are commonly used algorithms in moving object detection, including the difference method (background difference method and time difference method) and optical flow method and neural network. The difference method was based on the current video and the reference image subtraction to complete the detection. Some practical details for creating the Neural Network and image recognition in the Caffe Framework are given as well.

Keywords: *Detection of moving objects; tracking of moving objects; behavior understanding, Neural Network, Caffe model, CNN.*

I. INTRODUCTION

The exponential growth in hardware facilities like cameras, processing machines, mobile phones have led to an explosion of studies in automated video analysis for object detection and tracking. It is one of the hottest topics of research in computer vision and image processing. Object detection and tracking in video sequence is the key technology in the development of various video analysis applications that tires to detect and track objects over a sequence of images by replacing old traditional methods of monitoring cameras by human operators. The proposed solutions range from low cost handheld devices or cameras to high cost sophisticated and proprietary solutions. Object detection is the process of locating the occurrence of object using number of techniques like background subtraction, feature extraction, statistical methods etc.

Object detection is associated with object localization and tracking. This is evident through the creation of several object detection models such as YOLO (You Look Only Once) [1], SSD (Single Shot Detector) [2], and Faster-RCNN (Region-based Convolutional Neural Network) [3]. The current implementations, which concern the models mentioned above, utilize a type of deep learning network known as Convolutional Neural Networks (CNNs). These networks utilize a system of layers that perform convolutions on input images with several masks of varying kernel sizes. The objective of these convolutions is to obtain various grades of features from the images, and to utilize them as the key identifiers that discriminate a specific object from others. An approach that is commonly undertaken is to perform a combination of convolutions and sub-sampling to obtain the strongest and best features to represent certain objects, after which a series of fully connected networks would distinguish these features based on their different classes. An example is in the LeNet model [4] which performs a series of convolutions and pooling in the lower levels of the network, and incorporates a fully connected series of layers at the higher end of the network for plotting regression lines [5].

In this paper, we adopt SSD [2] to detect object from an image because it is one of the state-of-the-art algorithms which can achieve high detection accuracy in real-time. The SSD was developed by Wei Liu et al [2] as an improved version of Faster-RCNN and MultiBox [6]. It is unique from other detectors in that it introduces prediction of class confidence scores, together with an offset for the location of bounding boxes. SSD does so by using small 3 × 3 convolution filters on the feature maps, with a dedicated filter for each scale. This results in the ability of SSD to predict object classes from different aspect ratios as quickly and accurately as 58 Frames Per Second (FPS), 72.1% Mean Average Precision (MAP).

As mentioned above, the unique architecture of SSD has enabled it to have high class prediction accuracy, and high frame processing rate. SSD was developed over the VGG16's feed-forward convolution model. In order to study the SSD model, a Python script provided by Jia et al [7] is used to print the SSD model architecture. Figure 2 displays the basic structure of a series of layers from the base VGG16 model [8] that SSD was developed over.

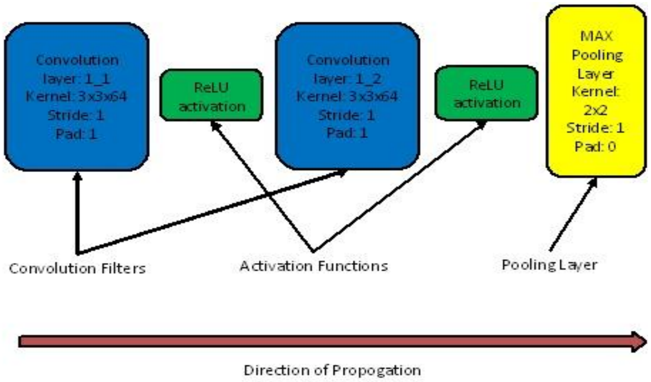

Convolution layer: 1_1 Kernel: 3x3x64 Stride: 1 Pad: 1
ReLU activation
Convolution layer: 1_2 Kernel: 3x3x64 Stride: 1 Pad: 1
ReLU activation
MAX Pooling Layer Kernel: 2x2 Stride: 1 Pad: 0
Convolution Filters
Activation Functions
Pooling Layer
Direction of Propogation

Figure 1.1: Basic structure of series of layers in VGG16.

With the base VGG16 network and the additional feature layers being established, the overall SSD model as introduced by Liu et al [2] is depicted in Figure 1.5.

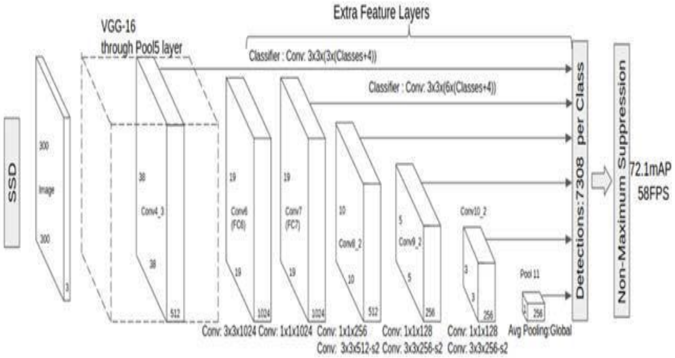


Figure 1.2: SSD300 model with breakdown of layers.

The SSD training objective [2] is derived from the Multi-Box objective but is extended to handle multiple object categories.

Motion detection is identifying varying regions from video frames using the fixed camera when objects are moving.

Difficulties in object detection using the fixed camera are:

(i) Objects movement in a scene from frame to frame do not have constant movement.

(ii) Objects in the scene may stop for some time and move further.

(iii) Objects in the scene have different velocities.

(iv)Moving objects in the scene may not cover a significant area of the frame this leads to recognition problem [9].

Substantial information about moving objects is required from frame to frame to track the objects properly. Deep Convolutional neural networks are best suited to do recognition and tracking of objects. Convolutional neural networks are the powerful visual model which has shown significant performance in many visual recognition problems. A Convolutional neural network is a combination of stacked Convolutional layers and spatial pooling layers which are stacked alternately. The Convolutional layer extracts feature maps using linear Convolutional filters and nonlinear activation functions. Spatial pooling performs grouping of the local features using spatially adjacent pixels, this improves the robustness towards objects deformation [10].

In Convolutional neural networks (CNN) original image can be used as input without pre-processing of the image. CNN has been showing the best accuracy in large-scale image classification/recognition since it is combined with deep learning [11]. Researchers optimize the CNN model structure by using parameters to improve the accuracy. Most of the improved models use more time to train and test [12]. CNN models involve large data, which is to say the training images cannot be too few. Considering the advantages of background subtraction based object detection and deep learning neural network based tracking is employed in this work.

II. RELATED WORK

In image classification analysis preprocessing is the essential step which must be evolutionary to remove the noises in the input image. The same principle is applied in the video processing and the detection of object is based on the video resolution and it makes the process simple to retrieve the necessary object form the frame sequence ad then tracking process becomes easier to implement in the real time scenario. Figure 2.1 has the general object detection phase which includes selecting a particular frame from the video and after selecting tracking section uses a object representation strategy to highlight the object or that particular portion then by using suitable dynamic model classification process is to be performed and in the last search mechanism enables the modules to keep tracking the object even in the crowd regions.

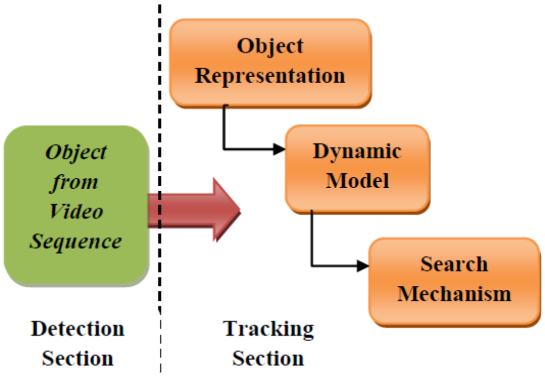


Figure 2.1: Object detection and tracking model.

Background subtraction process depends on the combination of features, models and classification, combination strategies. Most of the modelling techniques uses color intensity information as the key factor to create a edge information which is considered as an illumination. Edge information is not alone to be sufficient in identify the background since it has various intensities which are non uniform in nature. Second texture is used as a factor in background subtraction process it is a invariant function which is the combination of pixel and region based algorithms. Canny edge detection and haar transform in support to color intensities are some of the mixture combination available in texture bases classification. In some techniques information are completely different from background even the illumination is constant.

Detection of object or region is based on the movement in the region and background subtraction method is used to identify the portion. The process includes the following steps. Initialize the background for removal of noise using specific filters. In general median filter is used to remove the noises in the captured frame from the video segment. Second step is updating the background which is an updating process of moving object in real time and based on gray values moving objects are identified and justified using background values. In each step its values is updated frame by frame for all the sequences. Third step is a moving object extraction which has subtraction process for the actual image from the current image. In this process based on the threshold value it determines the pixel values and changes its present value if any variations present in it or else present value are remains in the same state as such and this process effectively suppresses the changes which occur due to lighting in the scenes. Reprocessing or difference determination [13] is the fourth step in the cyclic process which has noise removal step once again for the same image processes in step three. Many morphological methods are used further processing. In case of non human activity detection process filter process is essential to preserve the details. Fifth step includes extraction phase which extracts the frame from the background to enhance the details of the particular object. Shadow of the moving object is will affect the extraction based on the vertical and horizontal projection to detect the height and size of the object which is in movement. This is used to eliminate the shadow to a certain degree which is used to remove the pseudo local minimum value corresponds to moving object for précised edge. Hence different moving objects such as vehicles, trees and plants, floating clouds and other moving objects and the criteria is based on the two factors. First one is the object area which must be larger than the threshold value and aspect ratio of the object is based on the movement of the object to find the area and centroid of the object which is captured.

2.1 Video Target Tracking

Target tracking technology draws on cutting-edge technologies and advanced ideas in artificial intelligence, image processing, automatic control, and many other fields. The target tracking system is more intuitive, and the target's motion state can be directly observed in the video monitor, thus more detailed information about the target can be obtained. Particle filter algorithm is a relatively common algorithm in target tracking in recent years. It can be applied to nonlinear and non-Gaussian environments due to this superiority. Meanwhile, the particle filter's applications are more extensive than other algorithms, and have a lot of research space.

Video target tracking is mainly divided into several steps: target detection, feature extraction, identification and tracking, as shown in Fig.2.1.

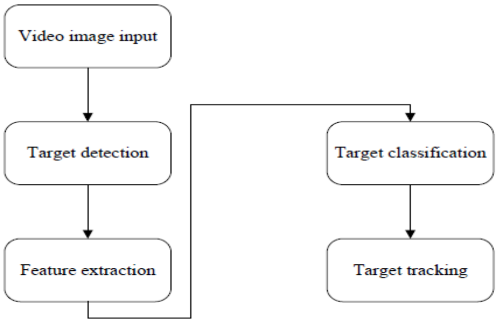


Figure 2.2: The process of video target tracking.

First detect the target in the video image, after that use the features of the target to identify the target, and finally use the tracking algorithm to track. So the detection of the target is basic in the video tracking. Target detection method commonly used inter-frame difference method, background difference method, etc. According to the characteristics of the target, various features of the target can be extracted such as: color features, texture features, edge features, and so on [14]. Recognition and tracking can be implemented using related algorithms such as particle filter algorithms.

The particle filter means: The probability density function of the state variable is approximated by a set of random samples propagating in the state space, and integral operation is replaced by the sample mean, finally obtain the minimum estimated variance of the state variable. These samples are called "particles," so called particle filters [15]. The basic principle of particle filtering is: Firstly, according to the distribution of state variables, a random set of samples is generated, and these samples become particles. Then according to the measured value to adjust the position of the particle and weights, modified the original distribution. Its essence is to use a set of randomly generated samples with sample weights to find the posterior probability density function of the state variables.

2.2 Moving object detection and tracking Using CNN

CNN is based learner because it is demonstrated to extract the local visual features and they are used in the recognition algorithms. CNNs require the extraction of local characteristics by limiting the receptive fields of the hidden units as local, based on the fact that the images have strong local two-dimensional structures. The convolutional neural network combines three architectural ideas to guarantee a certain degree of invariance of change and distortion: local receptive fields, shared weights (or pending replication) and sometimes, spatial and temporal sub sampling. Figure below shows the different types of ANN.

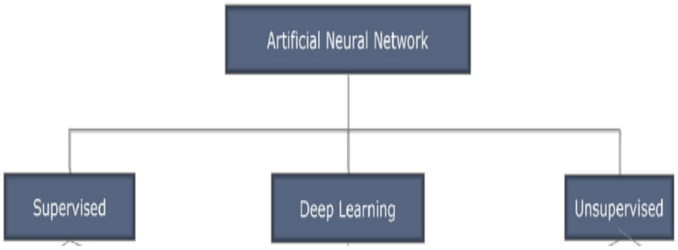


Figure 2.3: Types of ANN used for object detection.

Tracking objects is a fundamental problem in computer vision. Traditional methods based on feature, such as those based on color or blobs movement, follow-up maintaining a simple model of the objective and adapting this model over time. However, real situations in practice pose enormous challenges to these techniques because:

1) Over time, the model of the object can deviate from its original, and

2) Do not have a discriminating model that distinguishes the category of interest from the others.

There are different approaches had been presented by different researchers starting from background subtraction to CNN? Some of the human tracking methods have been presented in this section. Human tracking consist of three basic steps for pedestrian tracking: Human detection from sequence of frame, tracking and analysis of the tracking for particular purpose.

There are three fundamental aspects of pedestrian tracking that are analogous to object tracking:

1) Detection of the pedestrian in the video frame,

2) Tracking of the detection, and

3) Analysis of the tracks for the specified purpose.

In this literature survey, object feature point detection, background subtraction, segmentation and classification algorithms of previous research have been discussed. For tracking to be perfect, features which described the object is most important, hence the object detection is plays vital role. This can be achieved by using deterministic or probilistic motion models and appearance based model. To achieve the better accuracy adaptations of the model have been presented over time. The feature points were trained and update in the process of tracking. Only problem to track the object is that it requires large number of features which cannot be always be possible. Recently, the CNN is used to image classification and recognition to improve the significant performance. CNN is trained with millions of images of different classes. CNN are the learning method which exploits the spatial information of an image and learn the complex features automatically. CNN is intrusive to the variation of an input.

Fan et al. [16] presented the CNN based object tacking algorithm with shift variant architecture. In this algorithm, the features were learned during online process. The spatial and temporal features are considered using pair of images instead of single image.

Hong et al. [17] presented the approach where the output of the last layer of the pre-trained CNN module is cascade with the on-line SVM to learn discriminative appearance models. The tracking is performed using Bayesian network with target specify saliency map.

Wang et al. [18] used pre-trained CNN model for online tracking. The CNN is used after parameter tuning to adjust the appearance of the object in the scene and probability map is created to instead of creating labels. Wang and Yeung [19] create a stack de-noising auto encoder offline model which learned from the offline training. This model transfers the knowledge from offline model to track the online object.

III. PROPOSED METHODOLOGY

The Many problems in computer vision were saturating on their accuracy before a decade. However, with the rise of deep learning techniques, the accuracy of these problems drastically improved. One of the major problems was that of image classification, which is defined as predicting the class of the image. A slightly complicated problem is that of image localization, where the image contains a single object and the system should predict the class of the location of the object in the image (a bounding box around the object). The more complicated problem of object detection involves both classification and localization. In this case, the input to the system will be an image, and the output will be a bounding box corresponding to all the objects in the image, along with the class of object in each box. An overview of all these problems is depicted in Fig. 3.1.

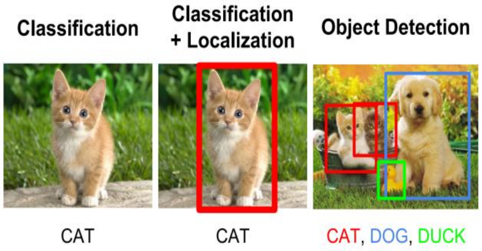


Figure 3.1: Object Detection.

Basic framework is developed on the principal of CNN. CNNs are particularly useful for finding patterns in images to recognize objects, faces, and scenes. They learn directly from image data, using patterns to classify images and eliminating the need for manual feature extraction. Using CNNs for deep learning has become increasingly popular due to three important factors: CNNs eliminate the need for manual feature extraction—the features are learned directly by the CNN. CNNs produce state-of-the-art recognition results.

CNNs can be retrained for new recognition tasks, enabling you to build on pre-existing networks.

3.1 Proposed Framework Details

Our proposed system usage Caffe (Convolutional Architecture for Fast Feature Embedding) framework for object detection. It is a deep learning framework, originally developed at UC Berkeley. Caffe provides a complete toolkit for training, testing, fine-tuning, and deploying models. CNN layers used in Caffe are as follows:

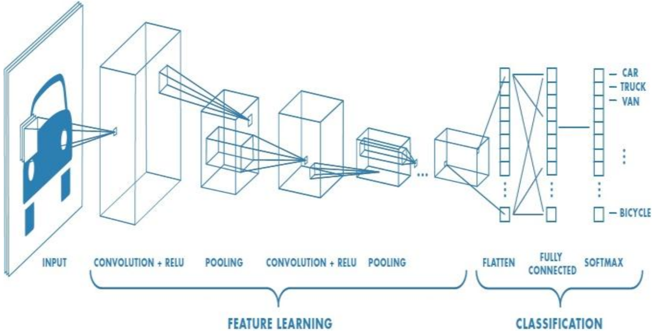


Figure 3.3: Detailed view with CNN Layers.

Convolution puts the input images through a set of convolutional filters, each of which activates certain features from the images.

Rectified linear unit (ReLU) allows for faster and more effective training by mapping negative values to zero and maintaining positive values. This is sometimes referred to as *activation*, because only the activated features are carried forward into the next layer.

Pooling simplifies the output by performing nonlinear down sampling, reducing the number of parameters that the network needs to learn.

These operations are repeated over tens or hundreds of layers, with each layer learning to identify different features. The proposed CNN based moving object detection algorithm consists of two phase: Object detection and tracking. The generalized block diagram of the proposed system is shown in Fig.3.4.

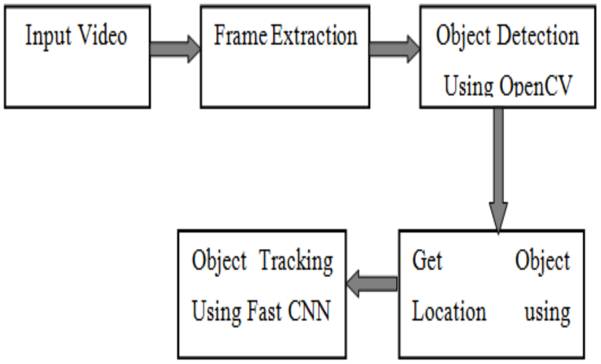

Input Video
Frame Extraction
Object Detection Using OpenCV
Get Object Location using
Object Tracking Using Fast CNN

Figure 3.1: Block Diagram of Proposed Model.

In this system, the video is feed to the system as an input. Frames are extracted for further processing. The two main algorithms object detection and object tracking is process through deep learning methods. The object detection is explained in detail in below flow. The object detection using computer vision algorithm is affected by different aspects like light variation, illumination, occlusion and system have difficulty to detect the multiple objects. Hence in this thesis, OpenCV based object detection algorithm has been used.

In this approach, firstly the necessary libraries are imported. Then import the pre-trained object detection model. The weights are initializing along with box and OpenCV class. After learning features in many layers, the architecture of a CNN shifts to classification. The next-to-last layer is a fully connected layer that outputs a vector of K dimensions where K is the number of classes that the network will be able to predict. This vector contains the probabilities for each class of any image being classified. The final layer of the CNN architecture uses a classification layer such as softmax to provide the classification output.

3.2 Proposed Processing Steps

Following are steps in object detection using Caffe pre-trained model:

Step-1: import the necessary packages.

Step-2: construct the argument parse and parse the arguments.

Step-3: initialize the list of class labels MobileNet SSD was trained to detect, then generate a set of bounding box colors for each class.

Step-4: load our serialized model from disk.

Step-5: initialize the video stream, allow the cammera sensor to warmup and initialize the FPS counter.

Step-6: loop over the frames from the video stream

(a) grab the frame from the threaded video stream and resize it to have a maximum width of 400 pixels.

(b) grab the frame dimensions and convert it to a blob.

(c) pass the blob through the network and obtain the detections and predictions.

```
(d) loop over the detections.
        -extract the confidence (i.e., probability) associated with the prediction
        -filter out weak detections by ensuring the `confidence` is greater than the
                    minimum confidence
        -if confidence > args["confidence"]:
            -extract the index of the class label from the`detections`, then compute
     the (x, y)-coordinates of the bounding box for the object
            -draw the prediction on the frame
(e) show the output frame
(f) if the `q` key was pressed, break from the loop
          -update the FPS counter
          -stop the timer and display FPS information.
```

IV. Results and Discussion

We applied the proposed object tracking method to the modified video thus produced resulting trajectories for moving objects in the modified video for two cases. In the first case the feature point's recovery was not implemented and in the second case the feature point's recovery was implemented according to the proposed method of regeneration.

A moving object was considered to be successfully tracked if the distance of its position in the resulting trajectory from its position in the reference trajectory does not exceed the specified maximum distance threshold for each video frame. The maximum distance threshold was specified manually in percents of the object bounding box size. The percentage of successfully tracked objects was computed as the number of successfully tracked objects divided by the total number of reference trajectories. The tracking deviation was computed as an average standard deviation of the resulting trajectories from the reference trajectories among all successfully tracked objects.

4.1 Caffe Framework

For the implementation of the proposed method, we used Caffe deep learning library. The main advantage of Caffe is the speed of operation. The framework supports CUDA and, if necessary, can switch the processing flow between the processor and the graphics card. The process of training the Deep Neural Network in framework Caffe has been lasted 30 epochs and finished with achieving the given accuracy of learning.

Figure 5.1 below shows the Caffe framework working model.

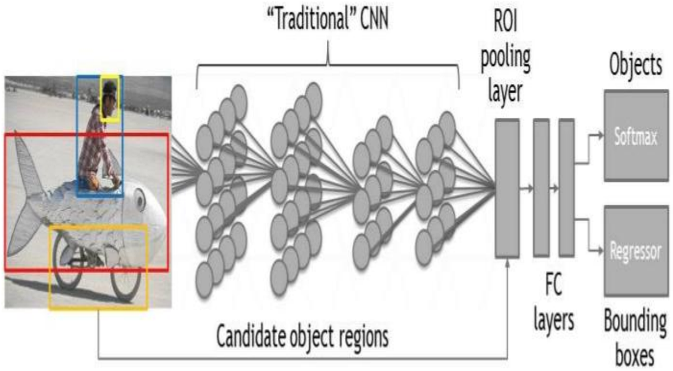
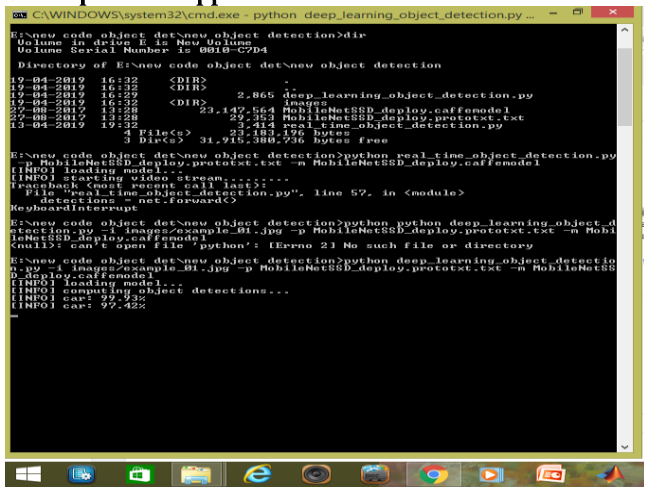


Figure 5.2: Region Based Caffe model for CNN.

4.2 Snapshot of Application


C:\WINDOWS\system32\cmd.exe - python deep_learning_object_detection.py ...

```
E:\new code object det\new object detection>dir
 Volume in drive E is New Volume
 Volume Serial Number is 0010-C7D4

 Directory of E:\new code object det\new object detection

19-04-2019  16:32    <DIR>          .
19-04-2019  16:32    <DIR>          ..
19-04-2019  16:29             2,865 deep_learning_object_detection.py
19-04-2019  16:32    <DIR>          images
27-08-2017  13:28        23,147,564 MobileNetSSD_deploy.caffemodel
27-08-2017  13:28            29,353 MobileNetSSD_deploy.prototxt.txt
13-04-2019  19:32             3,414 real_time_object_detection.py
               4 File(s)     23,183,196 bytes
               3 Dir(s)  31,915,380,736 bytes free

E:\new code object det\new object detection>python real_time_object_detection.py
 -p MobileNetSSD_deploy.prototxt.txt -m MobileNetSSD_deploy.caffemodel
[INFO] loading model...
[INFO] starting video stream.........
Traceback (most recent call last):
  File "real_time_object_detection.py", line 57, in <module>
    detections = net.forward()
KeyboardInterrupt

E:\new code object det\new object detection>python python deep_learning_object_d
etection.py -i images/example_01.jpg -p MobileNetSSD_deploy.prototxt.txt -m Mobi
leNetSSD_deploy.caffemodel
(null): can't open file 'python': [Errno 2] No such file or directory

E:\new code object det\new object detection>python deep_learning_object_detectio
n.py -i images/example_01.jpg -p MobileNetSSD_deploy.prototxt.txt -m MobileNetSS
D_deploy.caffemodel
[INFO] loading model...
[INFO] computing object detections...
[INFO] car: 99.93%
[INFO] car: 97.42%
_
```

Figure 5.3: Running existing algorithm using command line.

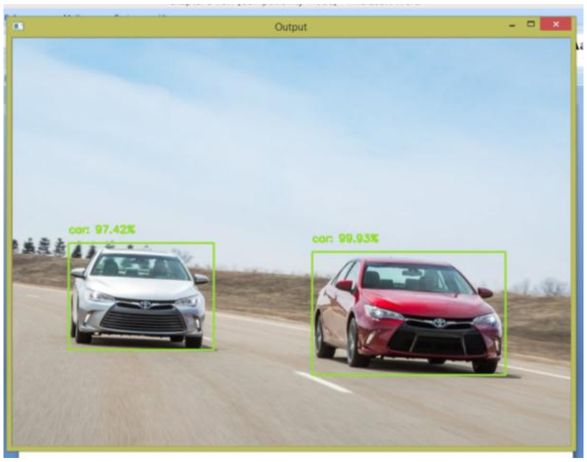


Figure 5.4: Output of existing method with accuracy of detection.

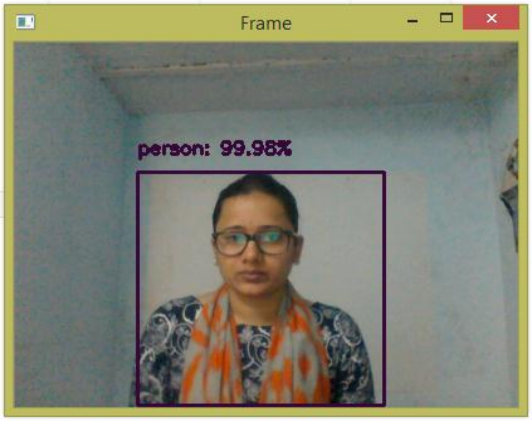


Figure 5.12: Output of proposed method with accuracy of detection of single image.

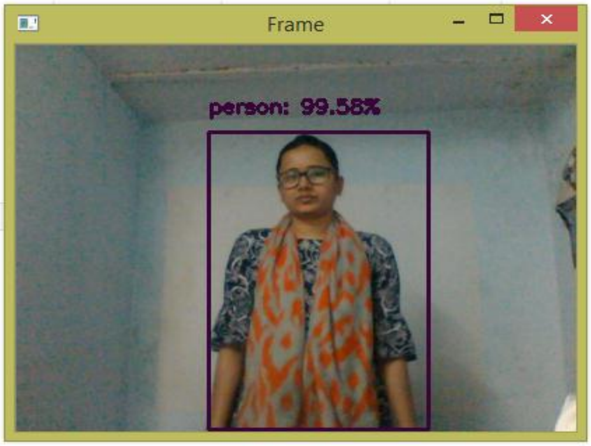


Figure 5.13: Output of proposed method with accuracy of detection of single image with increased distance.

Table below shows the comparison of existing system with proposed system. The percentage shows the accuracy of objects detection.

Table 4.1: Accuracy of Detection.

Object	Existing Method (%)	Proposed Method (%)
Person	88.44	99.98
Chair	65.03	85.05
Bottle	69.21	73.42

From above table we can conclude that accuracy of proposed moving object detection is high as compared to existing method.

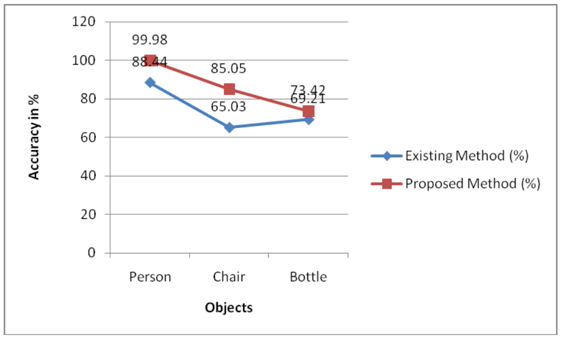


Figure 5.18: Comparision of existing and Proposed Method.

We performed the detection algorithm on all images in the system, and achieved near about 99% accuracy with respect to detection. Thus, our detection algorithm proved to be effective and fast as compared to existing algorithm.

V CONCLUSION

In this paper, novel approach for object detection and tracking has been presented using convolutional neural network. The moving object detection is performed using TensorFlow object detection API. The object detection module robustly detects the object. The detected object is tracked using CNN algorithm. Considering human tracking as a special case of detection of objects, spatial and temporal classes the facilities were learned during offline training. The shift variant architecture has extended the use of conventional CNNs and combined the global features and local characteristics in a natural way. The proposed approach achieves the accuracy of 95.85% to 99.25%.

The comparative results prove that proposed model improved the overall detection accuracy and as compared to traditional existing techniques for object detection.

REFERENCES

[1] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," IEEE CVPR, May 2016.

[2] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, A. C. Berg, "SSD: Single Shot Multi-Box Detector," https://arxiv.org/abs/1512.02325.

[3] S. Ren, K. He, R. Girshick and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," IEEE CVPR, Jan 2016.

[4] "Convolutional Neural Networks (LeNet)," Deeplearning.net, 2008. [Online]. Available: http://deeplearning.net/tutorial/lenet.html.

[5] Y. Lecun, L. Bottou, Y. Bengio and P. Haffner, "Gradient-based Learning Applied to Document Recognition," Proceedings of the IEEE, 1998.

[6] D. Erhan, C. Szegedy, A. Toshev and D. Anguelov, "Scale Object Detection Using Deep Neural Networks," IEEE International Conference on Computer Vision and Pattern Recognition, 2014.

[7] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, T. Darrell, "Caffe: Convolutional Architecture for

Fast Feature Embedding," BVLC, 2014. [Online]. Available: http://caffe.berkeleyvision.org/.

[8] K. Simonyan, A. Zisserman, "Very deep Convolutional networks for large-scale image recognition," International Conference on Learning Representations, Apr 2015.

[9] Tekalp AM, Digital video processing. Prentice Hall, 1995, New Jersey. B.N. Subudhi, S Ghosh , P.K. Nanda and A. Ghosh, "Moving object detection using spatio-temporal multilayer compound Markov Random Field and histogram thresholding based change detection", Multimedia Tools and Applications, vol. 76(11), June 2017, pp. 13511–13543.

[10]Shiqi Yu, Sen Jia and Chunyan Xu, "Convolutional neural networks for hyper spectral image classification", Neuro-computing, vol. 219, Jan.2017, pp. 88-98.

[11]Tianming Liang, Xinzheng Xu and Pengcheng Xiao, "A new image classification method based on modified condensed nearest neighbor and Convolutional neural networks", Pattern Recognition Letters, vol. 94, July 2017, pp-105-111.

[12]X.X. Niu and C.Y. Suen, "A novel hybrid CNN–SVM classifier for recognizing handwritten digits", Pattern Recognition, vol. 45, 2012, pp. 1318-1325.

[13]Deepak Kumar Panda; Sukadev Meher, "Detection of Moving Objects Using Fuzzy Color Difference Histogram Based Background Subtraction", IEEE Signal Processing Letters, Vol.23, No.1, pp.45-49, 2016.

[14]NingLi, TongweiLu, Yanduo Zhang. Object tracking algorithm based on the color histogram probability distribution. International Conference on Graphic and Image Processing, 2018.

[15]Fu Sun, JianXin Song. Research on Parallel Particle Filtering Target Tracking Algorithm Based on Hadoop. 2015 5th International Conference on Computer Sciences and Automation Engineering (ICCSAE 2015), 2016.

[16]Fan, J., W. Xu, Y. Wu, and Y. Gong, "Human tracking using convolutional neural networks", IEEE Transactions on Neural Networks 21(10), 1610-1623, 2010.

[17]Hong, S., T. You, S. Kwak, and B. Han, "Online tracking by learning discriminative saliency map with convolutional neural network", arXiv preprint arXiv: 1502.06796.

[18]Wang, N., S. Li, A. Gupta, and D. Yeung, "Transferring rich feature hierarchies for robust visual tracking", Computing Research Repository abs/1501.04587, 2015.

[19]Wang N., S. Li, A. Gupta, and D. Yeung, "Transferring rich feature hierarchies for robust visual tracking". Computing Research Repository 2015, abs/1501.04587.

AUTHORS PROFILE

Ms. Anjali Nema is pursuing Master of Engineering in CSE branch from Shri Ram Institute Of Technology, Jabalpur, M.P.

Anshul Khurana is currently working as a Asst. Prof. In CSE deptt. SRIT ,Jabalpur,M.P. He has done B.Tech and M.tech. in CSE branch.