Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 171 (2020) 2610–2617

Procedia Computer Science

www.elsevier.com/locate/procedia

Third International Conference on Computing and Network Communications (CoCoNet'19)

Object Detection System Based on Convolution Neural Networks Using Single Shot Multi-Box Detector

Ashwani Kumar*, Sonam Srivastava[b]

[a]*Vardhaman College of Engineering, Kacharam, Shamshabad – 501 218, Hyderabad, Telangana, India*
[b]*Institute of Engineering and Technology,Sitapur Road-226021, Lucknow, Uttar Pradesh, India*

Abstract

In this paper we propose object detection technique to detect objects in real time on any device running the model and in any environment. Object detection and training is a vast, vibrant and yet inconclusive and complex area of the computer vision. In this proposed work, convolutional neural network are used to develop a model which is composed of multiple layers to classify the given objects into any of the defined classes. The proposed schemes then use multiple images to detect the objects and label them with their respective class label. These objects are detected by making use of higher resolution feature maps. This is possible because of the recent advancement in deep learning with image processing. These images can be from the video frames which are fed into the model. Our scheme uses separate filters with different default boxes to tackle the difference in aspect ratio and also used multi-scale feature maps for object detection. The training of the model takes place until the error rate is less. The trained model is used to test some sample images. To speed up the computational performance of object detection technique we have use single shot multi-box detector algorithm along with the help of architecture of faster region convolutional neural network. The accuracy in detecting the objects is checked by the different parameters like loss function (LP), mean average precision (mAP), frames per second (FPS).

Keywords: Single Shot Multi-box Detector (SSMD); Faster Region Convolutional Neural Networks (F-CNN); Loss Function;

* Corresponding author. Tel.: +08413 – 253335; fax: +08413 – 253482.
E-mail address: ashwani.kumarcse@gmail.com

10.1016/j.procs.2020.04.283

1. Introduction

When we view an image the object in it are recognized by our brain instantaneously, but the machines take a lot of time for training and testing to identify the objects. Machines cannot do this task easily. People are trying hard to solve this problem, but they are able to achieve 65% of accuracy only. It is so hard for the machines to categorize and recognize objects like humans. This is actually the difficult task of computer vision.

Identifying each object in a picture or scene with the help of computer/software called as object detection. Face detection, driver less cars, vehicle detection and few other technologies uses object detection. For object detection, artificial neurons are used in deep neural networks they are similar to humans composed of neurons and process forwarding. The three major methods widely adopted in this field are: You Only Look Once (YOLO), Single Shot Detector (SSD) and Faster Region CNN (F-RCNN) [1]. In the first section, the introduction of object detection and deep learning process are shown. Second section demonstrates related work. In third section, we have reviewed the existing techniques of object detection. Section four, put a light on the proposed model. Fifth section represents the experimental results. Finally, section six concludes the proposed research work.

1.1. Our Contribution

In this section, author has provided some key point of the proposed object detection based on single shot multi-box detector algorithm.

1. The small objects are detected by making use of higher resolution feature maps. These higher resolution feature maps are taken from the higher resolution layers.
2. Low-resolution feature maps contains features like boundaries or edges and patches, that contains very small informative for classification. That is why we have ignored the low resolution layers.
3. In our approach we have also eliminated region of the image by doing this we can focus only in region of interest which leads our algorithm to speeds up the process for detecting the objects.
4. To strengthens the accuracy our algorithm applies some improvements like multi-scale features and include more number of default boxes.
5. This improvement further pushes the high speed. Combination of Faster R-CNN with convolutional features achieves the real-time processing speed and great accuracy.
6. Small convolutional filters are used to predict object classes labels.

2. Related Work

The following are the early works based on object detection models. Wei Xiang et al. [2] focus on the single shot detection (SSD) which is considered as newest algorithms for detecting the object. Thus, it is broadly observed that this SSD algorithm has comparatively smaller amount of precision in identifying little objects as compared to bigger objects. This is since it does not pay heed to the context from out of the proposal boxes. The paper presents shorthand for single shot multi-box detector i.e. CSSD. Two variants of CSSD have been discussed in this paper. The demonstration results show how multi-scale context modeling significantly enhances the precision in detection.

Reagan L. Galvez et al. [3] have shown that classification and detection of objects is now accurately possible with the recent advancements in the field of deep neural networks in image processing. In this paper, the authors have used CNN to detect the objects in the live environment [4]. Outputs very clearly show that former model is ideal for applications in real-time, the reason being its speed. Chengcheng Ning et al. [5] focused on Single Shot Multi-Box Detector (SSD) as one of the fastest algorithms in the field of object detection. It makes use of a single convolutional neural network for detecting the objects in an image. Although SSD algorithm of object detection is fast but still a big gap has been observed when the comparison is done. The authors propose a technique to enhance algorithm for increasing its accuracy of classification without any affect on its speed. Nashwan Adnan Othman et al. [6] recommended OpenCV libraries and deep learning technique for recognition for real-time video and an object detection. The authors have utilized Raspberry Pi 3 for the implementation of this system which helps in monitoring and capturing the frames and thus detecting and recognizing the objects. The application of few enhancements in the proposed method such as multi scale features, default boxes and depth wise separable convolution allow the our

approach to achieve a higher accuracy in detecting and recognizing objects. Hui Eun Kim et al. [7] focus on data augmentation specific to particular domain. In this paper, there is remarkable improvement shown by using the proposed method which is particular to on road detection of objects and it upgrades the average accuracy by 30%. Maria Jones and Viola [8] suggested that integral image is the new image representation, which allows the detector to detect the objects faster. AdaBoost algorithm from a larger dataset which picks least features and gives extremely efficient classifiers. Cascade method used for complex classifiers combining which enables quickly discard image background regions.

Romdhani et al. reduced set vectors are computed from the original set vectors in this model making the rejections as early as possible. Fleuret and Geman Starting from coaching examples, we have a tendency to recursively notice larger and bigger arrangements that are "decomposable" which means the chance of an appointment showing decays slowly on an object with its size. Schneiderman and Kande detects face expression and ignores anything like buildings, trees and alternative elements of body. Sung and Poggio uses the means of a view-based face and non-face model it distributes the human face patterns [9].

3. Reviewing Existing System

The deep neural network are basically consist two different model first is convoluted and the other is non-linear relationships. In both model the object considered as a layered configuration of primitives. In history there are numerous architectures and algorithms for implementing the concept of deep learning these network includes belief network, stacked network, gated recurrent unit etc. The first CNN was constructed by LeCun et al. [10]. The different application domain of convoluted neural network are image-processing, handwriting character recognition etc. Object detection is performed by estimating the coordinates and class of a particular objects in the picture. The presence of these object in an picture may be in random positions [11]. In this section, we have discussed only faster RCNN and YOLO v3 architecture.

3.1. Faster RCNN

Region Proposal Network for generating regions and detecting objects uses two methods of Faster-RCNN [12]. The first method proposes regions and uses the proposed regions respectively. In faster R-CNN the author [13] has used 16 architecture in convolution layers to achieve detection and classification accuracy on datasets. Kumar, A. et al. proposed different buyer seller watermarking protocol to provide secure and private transaction between the communicating parties [14-15].

3.2. Yolo v3

It elaborates to for you only look once. This is a detector of objects which makes use of features learned by a deep convolutional neural network for detecting object in real time. It consists of 75 convolutional layers, with up-sampling layers and skip connections for the complete image one neural network is applied. Regions of the image are made. Later bounding boxes are displayed along with probabilities. The most noticeable feature of v3 is that the detections at three different scales can be done with the help of it. Ashwani Kumar et al. [16-19] proposed different buyer seller watermarking protocol to provide secure and private transaction between the communicating parties.

4. Proposed Approach

In this section, we have shown the proposed approach for detecting the objects in real-time from images by using convolutional neural network deep learning process for that we have used OpenCV libraries. The proposed scheme uses single shot multi-box detector (SSMBD) algorithm for higher detection precision with real-time speed. However, the single shot multi-box detector (SSMBD) algorithm is not appropriate to detect tiny objects, since it overlooks the context from out of the boxes. Our proposed approach uses a new architecture as a combination of Faster R-CNN with convolutional features and SSMBD with multi-scale contexts in additional layers. The algorithm

comprises of two phases: first is feature maps extraction and the other one is concerned with application of small convolutional filters for detecting objects. The major objective during the training is to get a high class confidence score by matching the default boxes with the ground truth boxes. Our scheme uses separate filters with different default boxes to tackle the difference in aspect ratio and also used multi-scale feature maps for object detection.

The advantage of having Multi-Box on multiple layers leads to significant results in detection. Single Shot Multi Box Detector was discharged at the tip of Gregorian calendar month 2016 and thus arrived at a new set of records on customary knowledge sets like Pascal VOC and COCO. The major problem with the previous methods was how to recover the fall in precision, for that SSMBD applies some improvements includes multi-scale feature map and default boxes. For detecting small object higher resolution, feature maps are used. The training set of SSMBD algorithm depends upon three main sections i.e. select the size of box, matching of boxes and loss function. The proposed scheme can be understand by the system model given in figure 1.

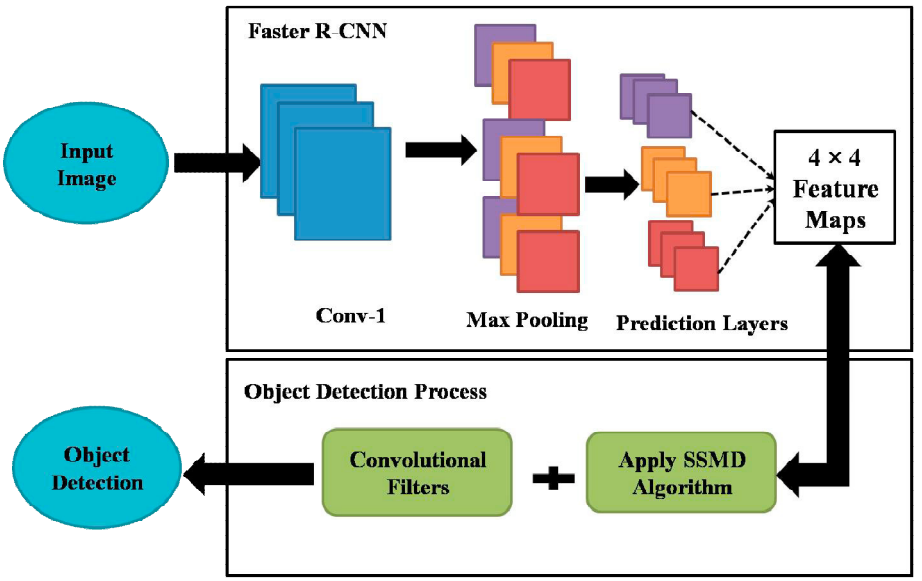

Fig. 1. The Proposed system model.

4.1. Steps in SSMBD Algorithm:

Algorithm 1: *Select the size of Box B*

Inputs:

I(x)←Input Image
C_l← Convolutional Layer
S(b)←Size of Box
$F_{(m)}$←Feature Map
d← dimension of boxes 4×4, 8×8, 16×16
$I_{(c)}$←Change in Intensity of pixel

Output:

B←2^d no. of boxes in Image

Procedure:

Initialize the size of box from 1 to d
for each *size of box S(b)identify feature map* **do**
a. $F_{(m)}$←{ Minimum C_l +Maximum $I_{(c)}$}
end for
for each $I_{(c)}$ *do*

if $I_{(c)}==1$ **then**
calculate
i. *Width (w)*= $C_l \times I_{(c}$
ii. *Height (H)*= $C_l \div I_{(c}$
else
resize the box with other possible dimension
end if
end for

4.2. Select the size of the boxes

The selection of boxes is based on the minimum value of convolution layer and maximum values of change in intensity [14]. SSMBD uses an extra box to make eight boxes for a specific map. The first Algorithm represent the procedure of producing specified feature maps F(m).

4.3. Identify the truth boxes

After finding the size of boxes, the next phase is matching of the boxes with the corresponding truth boxes. For a specific given picture to identify the truth Boxes is explained in the second algorithm.

4.4. Loss Function

The loss function is unbelievably simple and it is a methodology of evaluating how well your role models your dataset. If your predictions are entirely of your loss function can operate next range. If the output range is less, it means that the model is good. The main objective is to minimizing loss function. The loss function is also depends upon the sum of weighted localization and classification loss functions. Equation no.(a) represents the loss function [21].

4.5. Steps in identifying box size:

Algorithm 2: *Set of match box M(B)*

Inputs:
$\alpha \leftarrow$*Threshold value*
$t \leftarrow$ *No. of truth box*
$b \leftarrow$ *Number of default boxes*
$B \in 2^b$ *Set of boxes*
$T \in 2^{t\times 4}$ *Truth boxes set*
class[l]$\leftarrow$*Class labels set*
$N \leftarrow$ *Total no. of class labels*
Obj $\leftarrow$*Final Object*
Procedure:
Initialize the all object with default values
for each j^{th} *box B[i]* **do**
for each i^{th} *truth box having class label class[l]***do**
Match box (B[i],class[l])=1- F(m) $\leftarrow$*{ Minimum Cl +Maximum I(c)}*
if *(B[i],class[l])* $\geq \alpha$ **then**
class[l]=1
i. *Identified the Object (Obj)*
ii. *Label the Object (class[l])*
else

class[l]=0
Go *to step no. 2* **until** *the class label identified*
end if
end for
end for

Output: *Q* ← *Total no. of class labels*
I(p) ← *Indexed of positively boxes*

5. Result Analysis

We propose object detection technique to detect objects in real time on any device running the model and in any environment. The work is design for detecting the objects in real-time from images by using deep learning process. We have used Python programming language and Raspberry Pi 3 to execute the proposed system. Our proposed technique includes OpenCV 2.4 library. Python libraries are the open source framework for the construction, training and identification of object detection. The chosen dataset taken into consideration for this research was bound to a group of person. Multi-scale feature extraction may improve accuracy for detecting big object but does not exhibit a good precision of speed to detect small object.

For producing the result, we have used Pascal VOC and COCO datasets with ground truth bounding boxes and assigned class labels. Following operation is performed by the proposed model. Mean average precision (mAP) is used to measure the aaccuracy. The box regression technique of SSMBD is use to identify the bounding box coordinate. The accuracy is calculated using the below equation no. (1) and it could be improved over the original dataset also. It is used to measuring correctness of the classification process.

$$A_{ccuracy} = \frac{Object \quad (O_{correct})}{Total \quad Object \quad (T_{obj})} \tag{1}$$

Where $O_{correct}$ = number of correctly detected object and (T_{obj}) total number of images.

In the object detection techniques data augmentation plays a vital role to improve from 65.5% to 74.3% mAP. In the case of default box shapes, it improves from 71.6% to 74.3% mAP. Our SSMBD algorithm uses the 4× 4 feature maps. While comparing the other previous model the testing speed of proposed model is still fast because our approach gives 78 % of mAP and 89 FPS. Table 1 demonstrates the comparison between F-CNN, YOLO [16], SSD512, SSD300 and our proposed model i.e. F-CNN+SSMBD. SSD300 and SSD512 method performs Faster R-CNN in both speed and accuracy but not more than 70 % able to achieve of accuracy. Hence, we have combined Faster R-CNNT with SSMBD together to achieve high accuracy and FPS with good speed to detect objects real-time as well. Figure 2 demonstrate the different test images with object detection used in the experimental setup.

Fig. 2. Some sample images and object detecting F-CNN and SSMBD model.

Figure 3 represent the various object detected by the proposed algorithm in this we have used different colors of boxes to show different class labels. Our scheme correctly detects and recognizes bottle, laptop, mouse, cup, teddy bear, and umbrella, person, keyboard, TV, Zebra, toy car, bowl, chair, bird, vassal, suitcase. Table 1 represents the different parameter of the proposed method by using VOC and COCO test dataset.

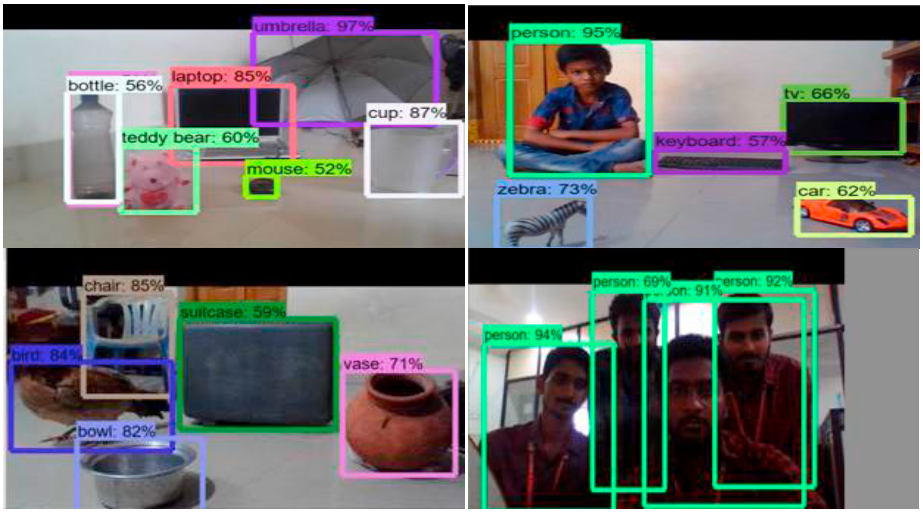

Fig. 3. Detection of object with different boxes using proposed approach.

Table 1 represents the results on Pascal VOC and COCO test.

System Model	mAP	FPS	No. of Boxes	Input Resolution
F-CNN	73.2	7	6000	1000×600
YOLO	66.4	155	98	448×448
SSD512	76.8	19	24564	512×512
SSD300	74.3	46	8732	300×300
F-CNN+SSBMD	78.68	89	5988	1024×1024

6. Conclusion

We have developed an object detector algorithm using deep learning neural networks for detecting the objects from the images. The research work uses a single shot multi-box detector (SSMBD) algorithm along with Faster CNN to achieve high accuracy in real time for detection of the objects. The performance of our algorithm is good in still images and videos. The accuracy of the proposed model is more than 75%. The training time for this model is about 5-6 hours. This model uses convolutional neural networks to extract feature information from the image and then perform feature mapping to classify the class label. Our scheme uses separate filters with different default boxes to tackle the difference in aspect ratio and also used multi-scale feature maps for object detection. The prime objective of our algorithm is to use truth box to extract feature maps. For checking the effectiveness of the scheme, we have use Pascal VOC and COCO dataset. We have compared the values of different metrics such as mAP and FPS with other previous model, which indicates that the algorithm achieves a higher mAP and uses more frames to gain good speed.

References

[1]. Redmon, J, Divvala, S, Girshick, R., Farhadi, A (2016) You only look once: Unified, real-time object detection. In: CVPR.

[2]. P. Poirson, P. Ammirato, C.-Y. Fu, J. Liu, Wei Koeck, A. C. Berg (2016) Fast single shot detection and pose estimation. International Conference on 3DVision(3DV).

[3]. R. L. Galvez, A. A. Bandala, E. P. Dadios, R. R. P. Vicerra, J. M. Z. Maningo (2018) Object Detection Using Convolutional Neural Networks. TENCON 2018-2018 IEEE Region 10 Conference.

[4]. D. Alamsyah and M. Fachrurrozi (2017) Faster R-CNN with Inception V2 for Fingertip Detection in Homogenous Background Image. IOP Conf. Series: Journal of Physics: Conf. Series 1196.

[5]. W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed (2015) SSD: Single shot multibox detector.

[6]. N.A. Othman, I. Aydin (2018) A new deep learning application based on movidius ncs for embedded object detection and recognition. 2nd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT).

[7]. Wei Xiang Dong-Qing Zhang Heather Yu Vassilis Athitsos (2018) Context-AwareSingle-ShotDetector. 2018 IEEE Winter Conference on Applications of Computer Vision, pp. 1784-1793.

[8]. P. Viola, M. Jones (2001) Rapid object detection using a boosted cascade of simple features. vol. 1, pp. 511-518.

[9]. C. Papageorgiou, M. Oren, T. Poggio (1998) A General Framework for Object Detection. Proc. IEEE Int'l Conf. Computer Vision.

[10]. LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521:1–10.

[11]. R. Girshick (2015) Fast R-CNN. in IEEE International Conference on Computer Vision (ICCV).

[12]. S. Ren, K. He, R. Girshick, and J. Sun. (2015) Faster r-cnn: Towards real-time object detection with region proposal networks. In Advances in neural information processing systems, pages 91–99.

[13]. Kumar, A., S. P. Ghrera, and Vipin Tyagi (2017) An ID-based Secure and Flexible Buyer-seller Watermarking Protocol for Copyright Protection. Pertanika Journal of Science & Technology 25.1.

[14]. Kumar, A (2019) Design of Secure Image Fusion Technique Using Cloud for Privacy-Preserving and Copyright Protection. International Journal of Cloud Applications and Computing (IJCAC) 9.3, 22-36.

[15]. Viral Thakar, Walid Ahmed, Mohammad M Soltani, Jia Yuan Yu (2018) Ensemble-based Adaptive Single-shot Multi-box Detector. in the Proceedings of the ISNCC 2018, 19-21 Rome, Italy.

[16]. Kumar, Ashwani, Satya Prakesh Ghrera, and Vipin Tyagi, “Modified Buyer Seller Watermarking Protocol based on Discrete Wavelet Transform and Principal Component Analysis”, Indian Journal of Science and Technology, 8(35), 1-9, 2015.

[17]. Kumar, Ashwani, Satya Prakesh Ghrera, & Vipin Tyagi, “Implementation of wavelet based modified buyer-seller watermarking protocol”, WSEAS Trans. Signal Process. 10, 212-220, 2014.

[18]. A. Kumar, S. P. Ghrera and V. Tyagi, "A new and efficient buyer-seller digital Watermarking protocol using identity based technique for copyright protection," 2015 Third International Conference on Image Information Processing (ICIIP), Waknaghat, 2015, pp. 531-535.

[19]. Kumar, Ashwani. "A Review on Implementation of Digital Image Watermarking Techniques Using LSB and DWT." Information and Communication Technology for Sustainable Development. Springer, Singapore, 595-602.