RESEARCH ARTICLE OPEN ACCESS

OBJECT DETECTION USING SINGLE SHOT MULTIBOX DETECTOR (SSD)

KAMAL CHANDWANI*, VEDITA JANBANDHU, ANJALI SELOKAR,
YASHKIRAN YERPUDE, KUNAL KAKDE, HASAN MESHRAM**
*(Assistant Professor, *Dept. Of Computer Technology, KDK College of Engineering, Nagpur*
Email: kamal.chandwani@kdkce.edu.in)
** (*Dept. Of Computer Technology, KDK College of Engineering, Nagpur*
Email: yashyerpude1999@gmail.com, hasanmeshram123@gmail.com, kunalkakde36@gmail.com,
veditajanbandhu02@gmail.com, anjaliselokar98@gmail.com)

Abstract:

In order to improve the detection accuracy of objects at different scales, the most recent studies applied multilayer architecture. However, the extracted low-level feature in the shallow layers may not work perfectly on the detection performance due to its less semantic information, especially for small objects. In this paper, we propose a refined feature-fusion structure to be integrated with single shot detector (SSD). To obtain the rich representation ability for feature mapping, in the fusion block, the deconvolution operation is basically applied to fuse high-level semantic features and low-level semantic features. It is noteworthy that in the proposed framework, the feature pyramid network is modified to better describe the features by the skip connection. An adaptive weighted connection is designed at the feature-fusion block, which further enhances the performance of the detection. On PASCAL VOC2007 test set, the experimental results show that the map of the proposed network is higher than SSD and deconvolutional single shot detector (DSSD) by 2.03% and 0.63%, respectively. Meanwhile, the speed of our method is as 2.2 times fast as the DSSD. Furthermore, the map of our refined feature-fusion structure SSD is 6.2% higher than SSD on the small object test set of PASCALS VOC2007, which verifies the effectiveness of the proposed model.

Keywords — **Python, CNN, Image processing, Datasets**

I. INTRODUCTION

The information age has witnessed the rapid development of wireless network technology, which has attracted the attention of researchers and practitioners due to its unique characteristics such as flexible structure and efficiency. As wireless network technology continues to evolve, it has brought great convenience to people’s life and work with its powerful technical capabilities. Wireless networks have gradually facilitated the mainstream of people’s online life. At the same time, the advent of 5G network will further enable the greater development and more advanced applications of wireless network technology. The future generations of wireless networks will provide strong support for related applications such as Internet of Things (IoT) and virtual reality (VR). Many of these applications connect to each other and transmit information within networks based on the detection of specific target objects. In order to achieve a comprehensive network connection between people and people, things and people, and things and things, one of the key tasks of future applications is to identify the target in a real-time manner in the wireless Identifying each object in a picture or scene with the help of computer/software is called object detection. Object detection is one of the most important problems in the area of wireless network computer vision. It is the basis of complex vision tasks such as target tracking and scene understanding and is widely used in wireless networks. The task of object detection is to determine whether there are objects belonging to the specified category in the image. If it exists, then the subsequent task is to identify its category and location information. Traditional object detection algorithms are mainly devoted to the detection of a few types of targets, such as pedestrian detection and infrared target detection. Due to the recent advance of deep learning technology, especially after the appearance of the deep convolution neural network (CNN) technology, object detection algorithms have made a breakthrough development. Within these algorithms, three major methods widely adopted in this field are You Only Look Once (YOLO), single shot multi-box detector (SSD), and faster region CNN (F-RCNN).

II. LITERATURE SURVEY

- For images or videos, many appliances may make use of fast and lightweight automated object detection. The technology industry has been continually introducing computational and hardware solutions, such as computers with impressive processing and storage capacities, over the past five years. Object detection methods, however, typically require either high processing power or large processing power, Availability of storage, which makes it difficult for resource-restricted devices to perform realtime detection without a strong server connection. In template matching, before a match is found, we slide a template image over a source image. But, since it has significant drawbacks, it is not the best tool for object recognition.

- That technique isn't durable. The following variables make the matching prototype a poor option for object detection.
  1. Rotation makes this technique inefficient.
  2. Size (known as scaling) also impacts this.
  3. Photo-metric alterations (e.g. brightness, contrast, hue etc.)
  4. Changes in the viewpoint of the distortion form (Affine).
- Human Based Translation of Neural Sign Language Estimation Key point:
  - There have been several ways of understanding hand languages that are used with a single hand to describe letters of the alphabet. As each letter of the alphabet simply corresponds to a particular hand shape, it is comparatively simpler than understanding sign languages. To identify the English alphabet with 92 percent recognition precision, the authors used depth cameras and the random forest algorithm. For understanding the American Sign Language, a pose estimation method of the upper body represented by seven main points was suggested (ASL).
  - We also note an approach by Kim et al. to identify the language of the Korean hand by examining latent hand picture characteristics. There are other ways by which signals are captured by using motion sensors like electromyography, for example (EMG) Sensors, RGB cameras, sensors from Kinect, and Controller of leap motion or their combinations.

- Hand gesture segmentation is a segmentation of the hand Done by using the technique of hand skin threshold. The device due to lighting, it does not yield successful outcomes.
- Conditions, interference with skin tone, and dynamic backgrounds increased noise. There are three kinds of skin colour that are International Machine Learning and Computing Journal.
- There are three kinds of skin colour that are Detection: the method of explicit range the nonparametric method process and the method of parametrises.
- The explicit variety the method divides the pixel class into skin- and skin- and from the allocated set of colours, non-skin-based

forms. This approach is primarily used because of its non-complex existence and reasonable computing pace.

III. PROPOSED WORK

Methodology

- This section presents our proposed approach for detecting the objects in real-time from images by using convolutional neural network deep learning process. The previous algorithms such as CNN, faster CNN, faster RCNN, YOLO, and SSD are only suitable for highly powerful computing machines, and they require a large amount of time to train. In this paper, we have tried to overcome the limitations of the SSD algorithm by introducing an improved SSD algorithm with some improvement The proposed scheme uses improved SSD algorithm for higher detection precision with real-time speed.

However, SSD algorithm is not appropriate to detect tiny objects, since it overlooks the context from the outside of the boxes. To address this issue, the proposed algorithm uses depth-wise separable convolution and spatial separable convolutions in their convolutional layers. Specifically, our proposed approach uses a new architecture as a combination of multilayer of convolutional neural network. The algorithm comprises of two phases. First, it reduces the feature maps extraction of spatial dimensions by using resolution multiplier. Second, it is designed with the application of small convolutional filters for detecting objects by using the best aspect ratio values. The major objective during the training is to get a high-class confidence score by matching the default boxes with the ground truth boxes. 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 which SSD applies some improvements that include multi-scale feature map and default boxes. For detecting a small object with higher resolutions, feature maps are used. The training set of improved SSD algorithm depends upon three main sections, i.e., selecting the size of box, matching of boxes, and loss function. The proposed scheme can be understood by the system.

★ **Pre-Processing Image**:

- These images were pre-processed with augmentation techniques in the pre-processing stage by adding pixels and colours, top and black hat, morphological transformation, blurring, saturation,
- and sharpening to enhance the variations in the datasets.
- The arbitrary size of the image in the datasets cannot be fed into the convolution neural network immediately.
- At the moment of reading images, the images were rescaled to 64 64 3 pixels by the deep learning library called TensorFlow during the model training.

represents the various objects detected by the proposed algorithm. In this research work, we have used different colours of boxes to show different class labels. Our scheme correctly detects and recognizes bottle, laptop, mouse, cup, teddy bear, umbrella, person, keyboard, TV, zebra, toy car, bowl, chair, bird, vassal, and suitcase.

★ **Recognizing pictures:**

- Using TensorFlow as the backend, OpenCV to read video frames, PyCharm code, and Python as the editor and programming language respectively, the qualified model was loaded onto a laptop.
- OpenCV captures video frames from the signer in Realtime and resizes them to 64 64 3 pixels.
- The model detects successfully. camera, the machine classifies the test sample and compares it with the gestures stored in the dictionary, and the corresponding output is shown to the user on the screen.

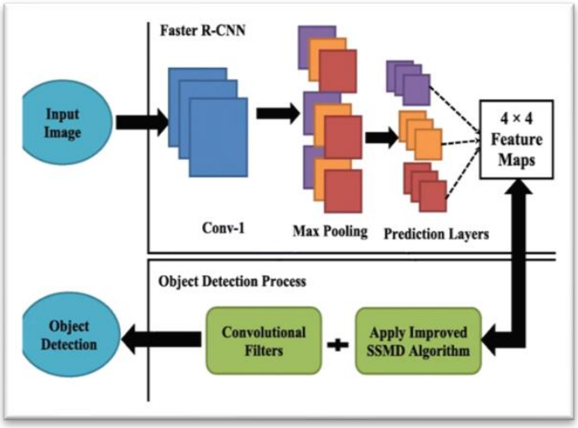

Fig 1: Conceptual framework diagram

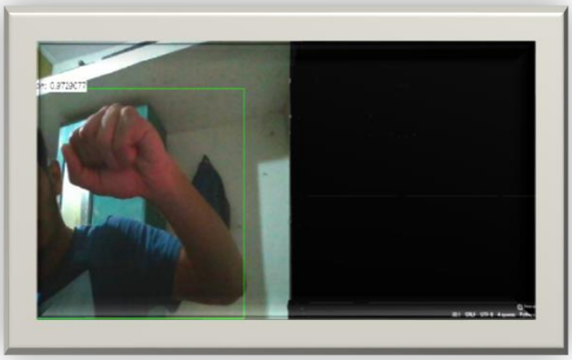

Fig 2. Recognizing video

IV. CONCLUSIONS

This study develops an object detector algorithm using deep learning neural networks for detecting the objects from the images. The research uses am improved SSD algorithm along with multilayer convolutional network to achieve high accuracy in real time for the 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 79.8%. The training time for this model is about 5–6 h. These convolutional neural networks extract feature information from the image and then perform feature mapping to classify the class label. The prime objective of our algorithm is to use the best aspect ratios values for selecting the default boxes so that we can improve SSD algorithm for detecting objects.

For checking the effectiveness of the scheme, we have used Pascal VOC and COCO datasets. We have compared the values of different metrics such as map, loss function, aspect ratio, and FPS with other previous models, which indicates that the proposed algorithm achieves a higher map, uses more frames to gain good speed, and obtains acceptable accuracy for detecting objects from colour images. This paper points out that the algorithm uses truth box to extract feature maps. Future research can extend our proposed algorithm by training the datasets for micro-objects

REFERENCES

[1] Y. Zhong, Y. Yang, X. Zhu, E. Dutkiewicz, Z. Zhou, T. Jiang, Device-free sensing for personnel detection in a foliage environment. IEEE Geoscience and Remote Sensing Letters **14**(6), 921–925 (2017). https://doi.org/10.1109/LGRS.2017.2687938

[2] S.Z. Su, S.Z. Li, S.Y. Chen, G.R. Cai, Y.D. Wu, A survey on pedestrian detection. DianziXuebao **40**(4), 814–820 (2012). https://doi.org/10.3969/j.issn.0372-2112.2012.04.031

[3] M. Zeng, J. Li, Z. Peng, The design of top-hat morphological filter and application to infrared target detection. Infrared Physics & Technology **48**(1), 67–76 (2006). https://doi.org/10.1016/j.infrared.2005.04.006

[4] W. Liu, D. Angelo, D. Erhan, C. Szeged, S. Reed, C.Y. Fu, A.C. Berg, SSD: Single shot multibox detector. In European conference on computer vision (pp. 21-37)(2016, October). Springer. Cham.. https://doi.org/10.1007/978-3-319-46448-0_2

[5]

Z. Li, F. Zhou, FSSD: feature fusion single shot multibox detector. arXiv preprint arXiv:1712.00960.(2017).

[6] J. Jeong, H. Park, N. Kwak, Enhancement of SSD by concatenating feature maps for object detection. arXiv preprint arXiv:1705.09587.(2017)

[7] J.Q. Wang,J.S. Li,X.W. Zhou,X. Zhang, Improved SSD algorithm and its performance analysis of small target detection in remote sensing images[J]. Acta Optica Sinica, 39(6):

0628005(2019).https://doi.org/10.3788/AOS201939.0628005.