International Journal for Innovative Engineering and Management Research

PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL

www.ijiemr.org

IJIEMR Transactions, online available on 10th Apr 2023. Link

:http://www.ijiemr.org/downloads.php?vol=Volume-12&issue=Issue 04

10.48047/IJIEMR/V12/ISSUE 04/118

Title **REAL TIME OBJECT DETECTION BASED ON SINGLE SHORT DETECTION ALGORITHM USING OPENCV FRAMEWORK**

Volume 12, ISSUE 04, Pages: 936-941

Paper Authors

Mr. B. Avinash, Rajesh Bolnedi, Boda Pradeep, Uma Maheshwar Sai Danda, Bhavanam Uday Eswara Reddy

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

Real Time Object Detection based on Single Short Detection Algorithm using OpenCV framework

Mr. B. Avinash, Assistant Professor, Department of IT,
Vasireddy Venkatadri Institute of Technology, Nambur, Guntur Dt., Andhra Pradesh.

Rajesh Bolnedi, **Boda Pradeep**, **Uma Maheshwar Sai Danda**, **Bhavanam Uday Eswara Reddy**
UG Students, Department of IT,
Vasireddy Venkatadri Institute of Technology, Nambur, Guntur Dt., Andhra Pradesh.
avinash.b@vvit.net, bolnedir@gmail.com, boda.pradeep596@gmail.com,
saidanda2001@gmail.com, udayeswarareddy@gmail.com

Abstract

In the area of computer vision, real-time object detection is a complex and dynamic task that includes localizing a single object or detecting multiple objects in an image. Modern deep learning models uses different networks for various tasks on embedded systems. In this project, a method has been developed for object recognition using the pre-trained deep learning model [1]. The objective is, to detect objects by capturing images from a webcam and subsequently detecting objects in a video stream which is specifically designed to be efficient for embedded systems [2]. To achieve the objective, the algorithm captures the images from the webcam and processes them in real-time using the Mobile Net model. The model processes the images and detects the objects present in them. This process is repeated for every frame of the video stream, enabling real-time object detection for the entire video. Overall, the developed method for real-time object detection using the Mobile Net model offers a promising solution for efficient and accurate object detection in various applications.

Keywords: Neural Networks, OpenCV, Object Detection, Mobile Net, SSD, Machine Learning

Introduction

Advancements in the field of deep learning applied to image processing have led to the emergence of innovative methods to detect and categorize objects in images by utilizing several input sources. These sources can be captured from video feeds, and the model is trained until the error rate becomes acceptable. To enhance the efficiency of detecting objects, we employed an optimizesd single-shot multi-box detector (SSD) algorithm and a faster region convolutional neural network. Multiple experiments were conducted with diverse parameters, including loss function, mean average precision (mAP), and frames per second, to assess the accuracy of our proposed method in object detection for real-time applications. The outcome of these experiments

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

illustrates that our technique has exceptional performance in correctly identifying and detecting objects.

By maintaining high accuracy, this system achieves 58 FPS with mAP 72.3% on MobileNet-SSD test, which is a substantial improvement over Faster R-CNN's [3] 7 FPS with mAP 71.8% or YOLO's [4] 45 FPS with mAP 64.5%. The increased speed is attributed to the use of a small convolutional filter for predicting object categories and bounding box locations, distinct predictors for different aspect ratio detections, and multiple layers for prediction at different scales. Overall, this new system significantly enhances real-time detection accuracy, achieving 74.3% mAP, which represents a larger relative improvement in detection accuracy than recent work on residual networks.

1.1 The Single Shot Detector (SSD)

The proposed SSD framework does the following

1. The image is processed through a multitude of convolutional layers that extract feature maps at various stages.

2. At every position within each feature map, a 4x4 filter is utilized to assess a default box that is small and low in size.

3. The model predicts the bounding box offset and class probabilities for each box.

4. The predicted boxes are matched with the truth boxes based on Intersection over Union rule.

5. Instead of using all negative examples, the model selects the best loss for each default box.

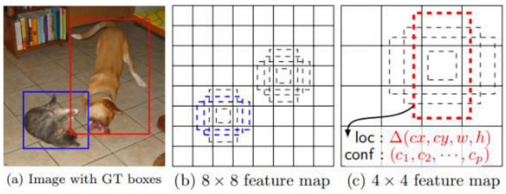

Fig:1 SSD Network

1.2 Model

The SSD approach utilizes a convolutional neural network to detect object class instances by producing fixed-size bounding boxes and corresponding scores. These detections are then refined through non-maximum suppression to eliminate redundancy and generate the final detections. Although the network has early layers that resemble a standard image classification architecture, it does not include any classification layers. To generate detections with specific key features, additional structure is added to the network. To detect objects at multiple scales, each feature layer employs a different convolutional model for predicting.

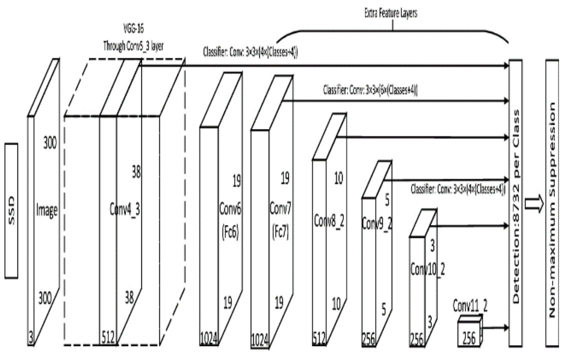

Fig:2 SSD Network Resolution

2. Literature Survey

Traditional object detection algorithms have multiple limitations, such as their high computational complexity, which requires sophisticated models and

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

extensive computation, making them unsuitable for real-time applications. They often require predefined regions of interest (ROIs) to be selected in the image, which can be time-consuming and may not capture all relevant objects in the scene. Additionally, these algorithms may struggle to accurately detect small or partially occluded objects, requiring retraining or fine-tuning for new object classes. Moreover, conventional object detection suffers from limited accuracy, lack of robustness to changes in object appearance, limited scalability, difficulty in detecting small objects, handling occlusion, detecting objects in cluttered scenes, and adapting to new objects without extensive training data. To address these limitations, Single Shot Detector (SSD) has emerged as a popular object detection technique in computer vision, enabling real-time object detection in both images and videos.

This literature survey reviews recent advancements and studies on SSD [5] for object detection, highlighting its benefits and drawbacks.

Advantages

1.SSD can perform object detection in a single shot, making it efficient and fast.

2.SSD can handle multiple object scales by using multiple layers with different receptive fields.

3.SSD is a simple and effective approach that achieves high accuracy on various object detection benchmarks.

4.SSD uses an end-to-end training approach, allowing it to optimize its parameters and detect objects simultaneously. This results in faster training times and improved accuracy.

5.SSD achieves high precision value in object detection tasks by utilizing feature maps at multiple resolutions, enabling it to detect objects of different scales.

6.SSD is superior at detecting small objects due to its use of feature maps with varying resolutions, which enables it to accurately detect small objects that traditional algorithms may miss.

Dis-advantages

1.SSD can suffer from class imbalance problems.

2.SSD can affect the detection accuracy.

3. Problem Identification

Real-time object detection has become an essential task in several fields, including surveillance, security, robotics, and autonomous driving. Accurately and quickly detecting objects in real-time has become a crucial aspect of many real-world applications. MobileNet-SSD is an ideal deep learning model for real-time object detection as it is specifically designed for mobile and embedded devices that have limited computational resources. The combination of MobileNet-SSD with Python and OpenCV can create a flexible and powerful object detection system that can work in real-time. Python is a user-friendly and versatile programming language, while OpenCV is a robust computer vision library with various features to analyze and process images and videos. The proposed project can help researchers and developers to enhance the accuracy and performance of object detection systems. It can also provide a valuable tool for organizations that require real-time object detection, including security companies, traffic control centers, and autonomous vehicle manufacturers. Moreover, the project can help advance the adoption of deep learning and computer vision technologies, as well as foster innovation and development in these areas.

4. Methodology

To develop a precise and resilient object detection model that can identify objects in complicated scenarios. The initial step is to acquire the COCO dataset [6], which includes over 330,000 annotated images of objects. The dataset is pre-processed by resizing the images to a standard size, converting them to grayscale, and normalizing the pixel values. SSD employs deep neural networks to identify objects in an image. The SSD model comprises a backbone network, feature extraction network, and detection network, with the backbone network being a pre-trained CNN, MobileNet. The feature extraction and detection networks are added to the backbone network.

The SSD model is then trained using the pre-processed COCO dataset, fine-tuning the pre-trained CNN (MobileNet) [7] to extract image features and training the detection network to predict object bounding boxes and labels. Once trained and evaluated, the SSD model can be used to detect objects in new images, returning predicted object bounding boxes and labels. Finally, the results obtained from the SSD model are analyzed, compared and illustrated using visualizations of the predicted bounding boxes and labels on sample images from the COCO dataset.

5. Implementation

The software and hardware requirements for object detection in images and video using using MobileNet SSD depend on various factors. Here are some general software and hardware requirements

5.1 Software Requirements:

1.Deep Learning Framework: You will need a deep learning framework like TensorFlow or PyTorch [8] to implement the MobileNet SSD model for object detection.

2.Object Detection Algorithm: MobileNet SSD as its backbone. You will need to download or implement this algorithm to perform object detection.

3.Video Processing Libraries: If you want to perform object detection in videos, you will need video processing libraries like OpenCV.

5.2 Hardware Requirements:

1. Graphics Processing Unit (GPU): Training a MobileNet SSD model requires a lot of computational power, so a GPU can speed up the training process. A high-end NVIDIA GPU like RTX 3070 or 3080 is recommended for training large models.

2. RAM: You need enough RAM to hold your dataset and to train your model. A minimum of 16 GB of RAM is recommended.

3. Storage: You need enough storage to store your dataset, model checkpoints, and other files. A minimum of 500 GB of storage is recommended.

4. CPU: While a powerful CPU is not mandatory, it can still help speed up some parts of the training process.

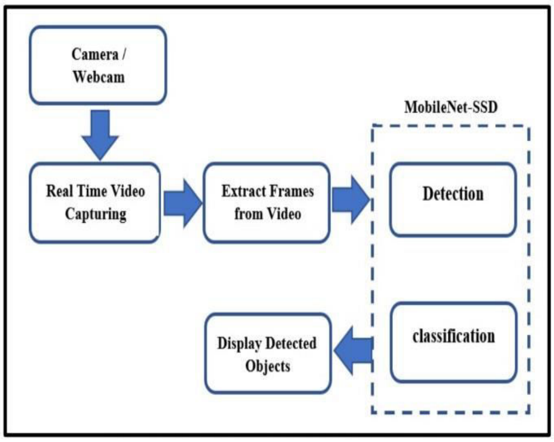

Fig:3 MobileNet-SSD Architecture

1.Open the real time webcam.

2.Extract frrames from video

3.For each frame in the video:

- Preprocess the frame by resizing it to the required input size and subtracting the mean and scaling the values.
- Pass the preprocessed frame through the MobileNet SSD model to obtain a set of predictions for objects in the frame.
- Filter out the predictions by detecting with low confidence scores.
- Draw the final set of bounding boxes on the frame.

4. Display the processed frame with the bounding boxes and accuracy overlaid on top.

6. Results & Conclusion

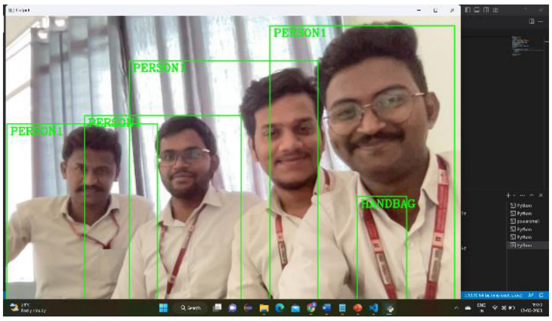

Fig 4: Detection in .jpg pictures.

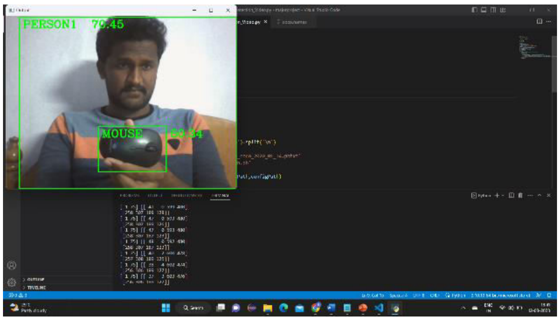

Fig 5: Detection with web cam

By using this SSD algorithm and based on experimental results, detecting objects in image and video is less amount of time when compared to other models. MobileNet SSD has demonstrated great potential for real-time object detection in various applications,providing valuable insights and enabling intelligent decision-making. Overall, object detection in images and videos using MobileNet SSD is a powerful tool for detecting the objects, and with continued research and development, it is likely to become even more effective and useful in the future.

7. Limitations

1. Limited accuracy: Although MobileNet SSD is good for small objects, it may not be as accurate as some of the more large complex models or server applications.

2. Limited ability to handle complex scenes: MobileNet SSD may struggle with

detecting objects in complex scenes that contain many overlapping objects or cluttered backgrounds. This can lead to false positives or missed detections.

Limited labelling: It only gives labels of the limited objects in live streaming video which are trained and tested in the COCO dataset.

8. Future Scope

1. Future research may focus on improving the speed and efficiency of MobileNet SSD for real-time video object detection, allowing it to be used in applications such as speed of autonomous vehicles.
2. For Nighttime visual tracking, night vision mode shoud be available as an inbuilt feature in the CCTV camera.
3. If climatic conditions are acceptable to the system, it can be encouraged to use to detect objects in underground mining factories.

9. References

[1] D. Erhan, C. Szegedy, A. Toshev, and D. Anguelov. object detection using deep neural networks. In CVPR, (2014).

[2] K. Karim, M. Akram, and M. H. Khan, "A survey on embedded systems and its applications," International Journal of Computer Applications, vol. 116, no. 10, pp. 1-5, 2015.

[3] Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: Towards real-time object detection with region proposal networks. In: NIPS. (2015).

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

[5] Liu, Wei, et al. "SSD: Single Shot MultiBox Detector." European Conference on Computer Vision (ECCV), 2016.

[6] Lin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., ... & Zitnick, C. L. (2014). Microsoft COCO: Common objects in context. In European conference on computer vision (pp. 740-755). Springer, Cham.

[7] Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., ... & Adam, H. (2017). MobileNets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861.

[8] PyTorch: Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., ... & Antiga, L. (2019). PyTorch: An imperative style, high-performance deep learning library. In Advances in Neural Information Processing Systems (pp. 8024-8035).