Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 5, May : 2023

A DEEP LEARNING APPROACH FOR EFFICIENT AND PRECISE OBJECT DETECTION

Harshal Batham, B.Tech., Dept. of Computer Science, Medi-Caps University.
Krish Gupta, B.Tech., Dept. of Computer Science, Medi-Caps University.
Dr Hemlata Patel, Professor, Dept. of Computer Science, Medi-Caps University.
Mr. Vivek Kumar Gupta, Professor, Dept. of Computer Science, Medi-Caps University.

Abstract
Deep learning has recently grown to be a significant influence in the artificial intelligence industry. Deep learning, a form of machine learning, involves training artificial neural networks with many layers to learn from and make predictions using huge, complex datasets. The most commonly employed techniques for object identification include Region-based Convolutional Neural Networks (RCNN), Faster-RCNN, Single Shot Detector (SSD), and You Only Look Once (YOLO).Faster-RCNN and YOLO are well known for their great accuracy whereas when deep learning MobileNet and SSD are combined, performance is not sacrificed for efficiency. The SSD-Mobilenet architecture is designed to be lightweight and optimized for mobile devices, making it faster and more efficient than Faster R-CNN and YOLO. Hence, SSD-MobileNet technique maintains excellent efficiency while successfully detecting objects.

Keywords: COCO, MobileNet, Single Shot detector, OpenCV, SSD-MobileNet.

I. Introduction

Object detection is a crucial element in computer vision and visual comprehension, playing a pivotal role in solving more advanced and high-level vision tasks like activity recognition, and object tracking. The introduction of deep learning has resulted in significant advancements in recent years, providing high levels of performance when applied to object detection [2,3,4]. Object detection has found practical applications in a diverse range of fields, including autonomous driving, medical, industrial, and surveillance scenarios, but there is still much untapped potential in this area.

The level of accuracy has not been adequate despite the usage of different object detecting algorithms. As a result, the neural networks were combined to find objects in video clips. [1]. Neural convolution networks are used in artificial vision to categorise images. This work and its companion papers discuss the development of object detection and tracking algorithms based on SSD and MobileNet in a Python environment. [1]. Finding the region of interest for an object belonging to a specific class inside an image is the goal of object detection. Frame differencing, optical flow, and background subtraction are a few techniques that can be used with a camera to locate and distinguish moving objects.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 5, May : 2023

II. Libraries, Algorithms and Dataset used

2.1 Single-Shot Detector(SSD)

By utilizing multibox, the Single Shot detector (SSD) approach may quickly recognize many objects in an image [5]. This model uses compression to transport default boxes between different feature maps and is based on the VGG-16 architecture. The object's shape is changed to fit its localized box, with shape offsets or confidence anticipated for each box, if the object seen belongs to the projected class of objects, and a score is generated. During the training phase, fully linked layers are removed, and ground truth containers and simple boxes are contrasted. The model loss, which measures the degree of discrepancy between expected and real boxes, is created by combining confidence loss and localized damage.

Confidence in the system is measured by the degree of confidence that the predicted object is the actual object. By eliminating feature re-sampling and encapsulating all computation in a single network, SSD is easy to train with MobileNet [2].

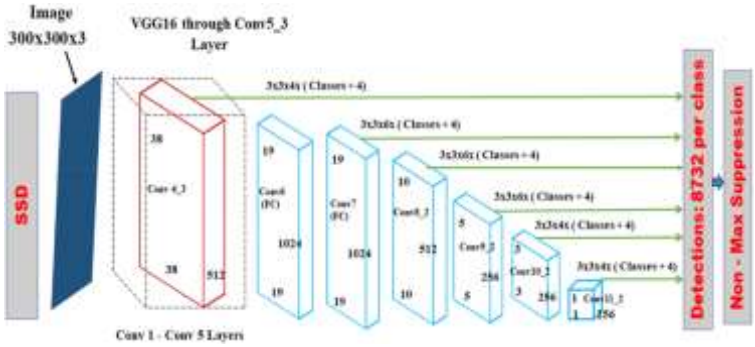


Figure 1: SSD [13].

2.2 MobileNet

Convolutional neural networks (CNNs) of the MobileNet class, which was open-sourced by Google [17], can be used to train extremely quick and compact classifiers. Deep neural networks can be produced by MobileNet, which use depth-separable convolutions. When process management is unavailable, mobile and embedded vision-based applications are especially well suited for the MobileNet architecture. When building tiny neural networks, MobileNet's main goal is to minimize latency. The design puts size before speed. These convolutions divide the input feature map into many feature maps after convolution, in contrast to typical convolutions [3].

The use of depth-wise separable convolutions in this model results in a significant reduction in the number of parameters, compared to networks with normal convolutions of the same depth. This parameter reduction leads to the creation of lightweight neural networks.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 5, May : 2023

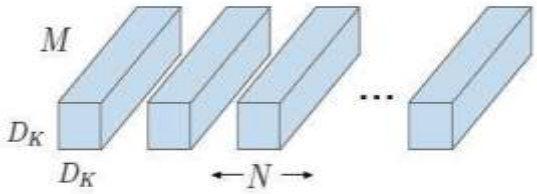


Figure 2: The regular convolutional filters are replaced by two layers [3].

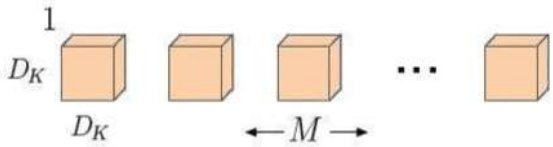


Figure 3: Depthwise convolution [3].

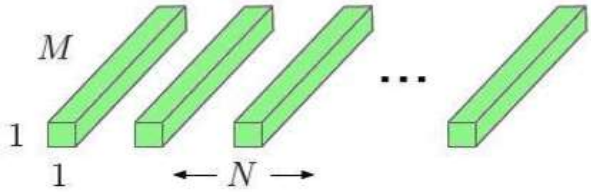


Figure 4: Depthwise Separable Convolution involves using 1x1 Convolutional Filters, which are also known as Pointwise Convolution [3].

2.3 SSD-MobileNet

MobileNet is a lightweight neural network architecture that can extract features from images while using fewer parameters, resulting in maintained performance. In SSD-MobileNet, the algorithm consists of two main components: the front-end contains the MobileNet network, which extracts the initial target characteristics, and the back-end consists of a multi-scale feature detection network that obtains typical features from the front-end network under various conditions. Information from six scales is directed towards the final detection module, which estimates the target's location, classification, and confidence [17]. To eliminate duplicate predictions, non-maximum suppression (NMS) module is used.

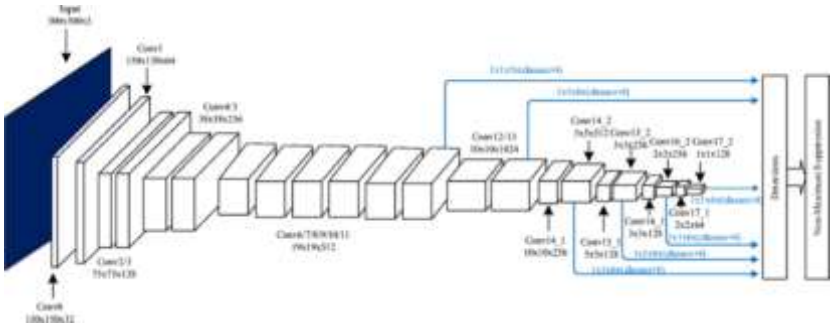


Figure 5: SSD+ MobileNet

2.4 Coco Dataset

A thorough tool for image classification, segmentation, and labeling is the COCO (Common Objects in Context) database, which has over 330,000 images and 2.5 million annotated object over 80 categories. In computer vision research, this dataset is commonly used as a reference for object detection systems. We used only 20 classes in this experiment that fulfill the object classification aim and maintain models training more reliable and acceptable [19].

III. Literature Review

According to references [6,7,8,9], single-stage object detection methods such as YOLO and SSD directly perform classification and detection from images using sliding windows of various sizes, thus eliminating the need for region proposal systems or post-classification layers. Although single-stage methods are generally faster than two-stage methods, they are typically less accurate. Many single-stage methods share a similar structure, comprising two sub-networks: a backbone network for feature extraction and a front-end network for object classification and localization. For instance, the SSD network [11] utilizes VGG16 as its backbone network, with a front-end network featuring multiscale convolutional layers to achieve fast detection and high quality.

One technique to enhance object detection performance is by using deeper networks, but doing so raises computational costs, making it challenging to employ on mobile devices with limited computing power. For classification tasks, however, the Mobilenet design [12] recommends depth-wise separable convolutions, which comprise a fully convolutional convolution followed by a point-wise convolution, considerably lowering the amount of operations required to compute convolutions. Many object identification models, like SSD Lite [13], have adopted Mobilenet as a feature extractor in place of models that employ conventional convolutions, which decreases their size and memory needs. Some models, like Tiny-DSOD [14], combine densely connected networks with complexity separable convolution operation to minimize model complexity. The Depth-wise Dense Block is an effective network structure for doing this.

IV. Methodology

The first stage entails creating an optical flow between the application and the capturing device, which in this example is a camera. To do this challenge, OpenCV is used which computer vision library is offering multiple functionalities [19]. The background of the visual stream must then be eliminated in order to accurately identify the object and determine its class. This forms the fundamental step of a multi-stage vision system, which differentiates the background from the foreground object in a sequential stream of images. The image's foreground or subject is recognised and distinguished from the background for additional pre-processing. The localization of the area of interest is done after a step-by-step demonstration of the separation process.

The project uses video stream as input, so there are frequently moving items in the movie. Thus, following the object's trajectory is a crucial next step. By accounting for the duration

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 5, May : 2023

between the item's displacements, the object identification process also has the potential to calculate the moving object's velocity. Nevertheless, as it adds complexity, this project does not prioritize figuring out the image scale in each frame for precise calculation.
After identifying the object, it generates a response to inform user. This is done by drawing boxes around detected objects that follow their movement. The detected class and confidence level are displayed above each object.

V. Result

The SSD-MobileNet approach was used to construct a program that was implemented in OpenCV. Twenty different objects were trained into the model. The following outcomes were attained when the recorded video from the camera was successfully scanned, detected, and tracked. Figure 7 illustrates the confidence level as it reached a very high percentage of 99.9%. A dog, motorcycle, person, potted plant, bird, automobile, cat, sofa, sheep, bottle, chair, aeroplane, train, bicycle, etc. are just a few of the 20 object types that the model was trained to recognize effectively. The program was written on a Jupyter Notebook incorporated in Anaconda environment. To avoid adding any unnecessary complications to the model, we selected only 20 classes for training purpose.
Furthermore, our system works on confidence levels and the graph in figure 6, shows the highest and lowest confidence levels detected during in the videos. The lowest confidence levels are capped at 0.5 because of specification of threshold confidence.

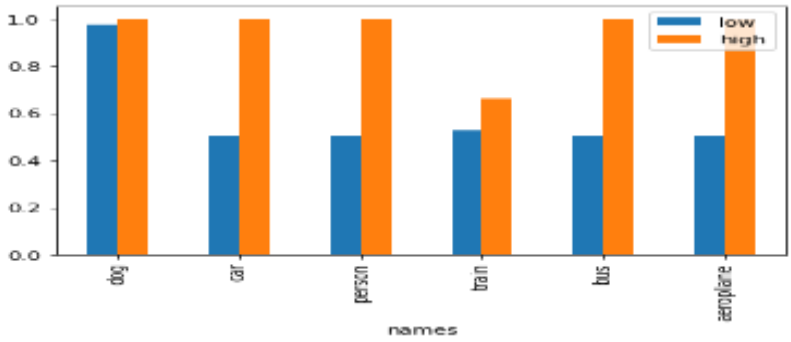
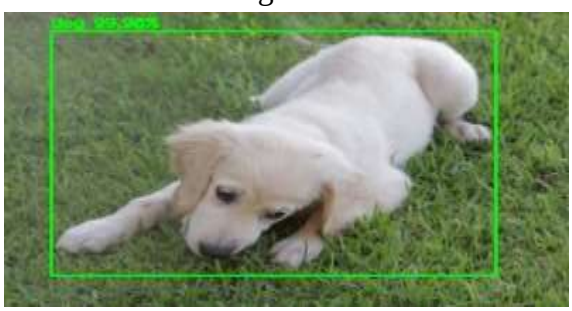


Figure 6



Class Predicted: Dog
Confidence: 99.98%

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 5, May : 2023

Class Predicted: Aeroplane
Confidence: 99.93%

Class Predicted: Bus
Confidence: 99.96%

Class Predicted: Person
Confidence: 100.00%
Figure 7

VI. Conclusion and Future Scope

Real-time object identification is done using SSD-MobileNet algorithm. This algorithm's primary goal is to find and follow several objects in a real-time video stream. The trained model generated good detection and tracking skills to find, follow, and react to the assessment and appropriate in the CCTV. The project can also be used for traffic control by utilizing public surveillance cameras to detect ambulance vehicles and adjust traffic signals accordingly. The technology can even be applied to satellite systems to prevent terrorist attacks by tracking their movements at the National borders. Furthermore, the utility of this system can be extended for the benefit of visually disabled people [9].

References

1. Kanimozhi, S., G. Gayathri, and T. Mala. "Multiple Real-time object identification using Single shot Multi-Box detection." *2019 International Conference on Computational Intelligence in Data Science (ICCIDS)*. IEEE, 2019.
2. Wei Liu and Alexander C. Berg, “SSD: Single Shot MultiBox Detector”, Google Inc., Dec 2016.
3. Andrew G. Howard, and Hartwig Adam, “MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications”, Google Inc., 17 Apr 2017.
4. Justin Lai, Sydney Maples, “Ammunition Detection: Developing a RealTime Gun Detection Classifier”, Stanford University, Feb 2017
5. Khandelwal, Renu. "SSD: Single Shot Detector for object detection using MultiBox." (2019).
6. J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only look once: Unified, real-time object detection,” in Proceedings of the IEEE conf. on computer vision and pattern recognition, 2016, pp. 779–788
7. W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, “Ssd: Single shot multibox detector,” in European conference on computer vision. Springer, 2016, pp. 21–37.
8. J. Redmon and A. Farhadi, “Yolov3: An incremental improvement,” CoRR, vol. abs/1804.02767, 2018.
9. Terreran, Matteo, et al. "Real-time object detection using deep learning for helping people with visual impairments." *2020 IEEE 4th International Conference on Image Processing, Applications and Systems (IPAS)*. IEEE, 2020.
10. Z. Shen, Z. Liu, J. Li, Y.-G. Jiang, Y. Chen, and X. Xue, “Dsod: Learning deeply supervised object detectors from scratch,” in Proceedings of the IEEE international conf. on computer vision, 2017, pp. 1919–1927
11. J. R. Uijlings, K. E. Van De Sande, T. Gevers, and A. W. Smeulders, “Selective search for object recognition,” International journal of computer vision, vol. 104, no. 2, pp. 154–171, 2013
12. A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, “Mobilenets: Efficient convolutional neural networks for mobile vision applications,” arXiv preprint arXiv:1704.04861, 2017.
13. Chintakindi Balaram Murthy Mohammad Farukh Hashmi Avinash G. Keskar “Optimized MobileNet+SSD: a real-time pedestrian detection on a low-end edge device” International Journal of Multimedia Information Retrieval https://doi.org/10.1007/s13735-021-00212-7.
14. M. B. Sandler, A. G. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, “Mobilenetv2: Inverted residuals and linear bottlenecks,” 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4510–4520, 2018.
15. Y. Li, J. Li, W. Lin, and J. Li, “Tiny-dsod: Lightweight object detection for resource-restricted usages,” in BMVC, 2018.
16. G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, “Densely connected convolutional networks,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 4700–4708.
17. G. Yu, L. Wang, M. Hou, Y. Liang and T. He, "An adaptive dead fish detection approach using SSD-MobileNet," 2020 Chinese Automation Congress (CAC), Shanghai, China, 2020, pp. 1973-1979, doi: 10.1109/CAC51589.2020.9326648.

18. PujaraAbhijeet, “Image Classification WithMobileNet”, 2023, https://builtin.com/machine-learning/mobilenet.
19. OpenCV about, OpenCV official documentation, https://opencv.org/about/.
20. Wang, Zhiwen, Jing Feng, and Yifeng Zhang. "Pedestrian detection in infrared image based on depth transfer learning." Multimedia Tools and Applications 81.27 (2022): 39655-39674.