Object Detection Using Deep Learning

Dinesh Suryavanshi (1) **Mr. Vijay Bisen (2)** **Mrs. Neha Khatri (3)**

Final year Student VITM Indore

HOD of VITM Indore

Professor VITM Indore

ABSTRACT

In today's scenario, the fastest algorithm which uses a single layer of convolutional network to detect the objects from the image is single shot multi-box detector (SSD) algorithm. This paper studies object detection techniques to detect objects in real time on any device running the proposed model in any environment. In this paper, we have increased the classification accuracy of detecting objects by improving the SSD algorithm while keeping the speed constant. These improvements have been done in their convolutional layers, by using depth-wise separable convolution along with spatial separable convolutions generally called multilayer convolutional neural networks. The proposed method uses these multilayer convolutional neural networks to develop a system model which consists of multilayers to classify the given objects into any of the defined classes. The schemes then use multiple images and detect the objects from these images, labeling them with their respective class label. To speed up the computational performance, the proposed algorithm is applied along with the multilayer convolutional neural network which uses a larger number of default boxes and results in more accurate detection. The accuracy in detecting the objects is checked by different parameters such as loss function, frames per second (FPS), mean average precision (mAP), and aspect ratio. Experimental results confirm that our proposed improved SSD algorithm has high accuracy.

Keyword - Object tracking, object detection, deep learning, Machine learning, SSD (Single shot detection), YOLO (You only look once).

INTRODUCTION

The study of computer programs that use algorithms and statistical models to learn through inference and patterns without being explicitly programmed is referred to as machine learning. In the past ten years, the field of machine learning has made considerable advancements. Machine Learning (ML) has proven to be one of the most game-changing technological advancements of the past decade. In the increasingly competitive corporate world, ML is enabling companies to fast-track digital transformation and move into an age of automation. Some might even argue that AI/ML is required to stay relevant in some verticals, such as digital payments and fraud detection in banking or product recommendations. A wide range of businesses have adopted machine learning at scale across sectors, and their eventual acceptance of these algorithms and their pervasiveness in enterprises are well-documented.

Machine learning is now used in some capacity by virtually every other piece of software and software application on the Internet. Machine learning is currently used by businesses to solve a wide range of issues since it has grown so prevalent.

LITERATURE SURVEY

1) OBJECT DETECTION -Traditional object detection and recognition systems have been around since 1990s. The typical approach here is to extract features using SIF or SURF and then use them to train a classification model like SVM or ANN. However, more recent advancements in deep learning have potentially replaced the previous approaches, providing increased accuracy with low effort.

Automatic Object Tracking and Segmentation Using Unsupervised Siam Mask

Received July 4, 2021, accepted July 21, 2021, date of publication July 28, 2021, date of current version August 5, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3101054

2) According to the actual detection performance, when the input image resolution is 300×300, the size of the last feature map is 1×1, so the accuracy and positioning of small object detection is not high. According to the characteristics of small object such as lower solution and complex background, the feature of small object in low-level network are more obvious. After several times of pooling of neural network, the feature maps of small

object will appear fuzzy, causing bad detection results. Therefore, we change the resolution of the input image from 300×300 to 428×428, and the network is optimized on SSD baseline.

Small Object Recognition Algorithm of Grain Pests Based on SSD Feature Fusion

ZONGWANG LYU, HUIFANG JIN , TONG ZHEN, FUYAN SUN, AND HUI XU

College of Information Science and Engineering, Henan University of Technology, Zhengzhou 450001, China.

Received February 23, 2021, accepted March 10, 2021, date of publication March 17, 2021, date of current version March 23, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3066510

3) MULTI-TASK LEARNING - In multi-task learning, a neural network can be trained to solve multiple learning tasks by exploiting commonalities and differences across tasks. In Uber Net, a multi-task network that tackles seven computer vision tasks is proposed by linking task-specific networks to a shared backbone network. In a gradient normalization algorithm is proposed to balance the training process for multi-task networks. In a cross-stitch unit is proposed to combine multiple networks to learn the best combination of shared and task-specific feature maps. In, the authors investigated if tasks should be learned together in a multi-task frame work so that the overall performance is optimized.

Improving Object Detection Using Weakly-Annotated Auxiliary Multi-Label Segmentation

ZHENGYU XIA , CHEN ZHANG , AND JOOHEE KIM , (Senior Member, IEEE) Illinois Institute of Technology, Chicago, IL 60616, USA Corresponding author: Zhengyu Xia (zxia@hawk.iit.edu)

Received October 20, 2021, accepted November 28, 2021, date of publication December 1, 2021, date of current version December 13, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3132225

4) DETECTION NETWORK SSAM-YOLO is proposed as a Yolo-based detection network, as shown in Figure 2. The backbone, neck, and head are the three main components of YOLO-based detectors. Due to their higher resolution and more accurate spatial features, the extracted feature maps of the backbone are more effective for vehicle detection than for other feature maps in the detection network. The head and neck are more helpful in classifying vehicles since they provide higher semantic data and depth, despite lower spatial detail due to lower resolution. This paper proposes a design for the SSAM-YOLO detector that improves the accuracy of vehicle position detection and increases scale change robustness in light of the MRF block and SSAM module in the backbone, referred to as the Semantic Attention Network (SemAtt - Net).

Fast-Yolo-Rec: Incorporating Yolo-Base Detection and Recurrent-Base Prediction Networks for Fast Vehicle Detection in Consecutive Images

NAFISEH ZAREI1, PAYMAN MOALLEM 1, AND MOHAMMADREZA SHAMS2 1Department of Electrical Engineering, Faculty of Engineering, University of Isfahan, Isfahan 81746-73441, Iran

2Department of Computer Engineering, Shahreza Campus, University of Isfahan, Isfahan 81746-73441, Iran

Corresponding author: Payman Moallem (p_moallem@eng.ui.ac.ir)

Received 25 October 2022, accepted 4 November 2022, date of publication 14 November 2022, date of current version 21 November 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3221942

5) In this paper, author consider that the deep learning object detection application is implemented on an edge device equipped with multi-core CPUs and GPUs. By exploiting the parallel computing capability of the GPUs, it is possible to accelerate the deep learning inference. In this section, we introduce a practical approach to optimize the deep learning object detection applications in edge computing environment.

A Method for Optimizing Deep Learning Object Detection in Edge Computing

Ryangsoo Kim, Geonyong Kim, Heedo Kim, Giha Yoon, and Hark Yoo Honam Research Center Electronics and Telecommunications Research Institute (ETRI) Gwangju, Republic of Korea {rskim, gykim, globet, gya, harkyoo}@etri.re.kr

RESEARCH GAP AND OBJECTIVE

Based on the literature we have identified few research gaps mentioned below.

1. The Yolo method is still in the exploring stage to identify and label the objects.

   The SSD method required more time to fetch the label and identification of the object.

2. The list of objectives are mentioned below,

   To study the YOLO method and explore the possibilities of labeling in a dynamic range on real time.

   To study the SSD method and explore the time reduction method for labeling.

PROPOSED METHODOLOGY

Object detection algorithms

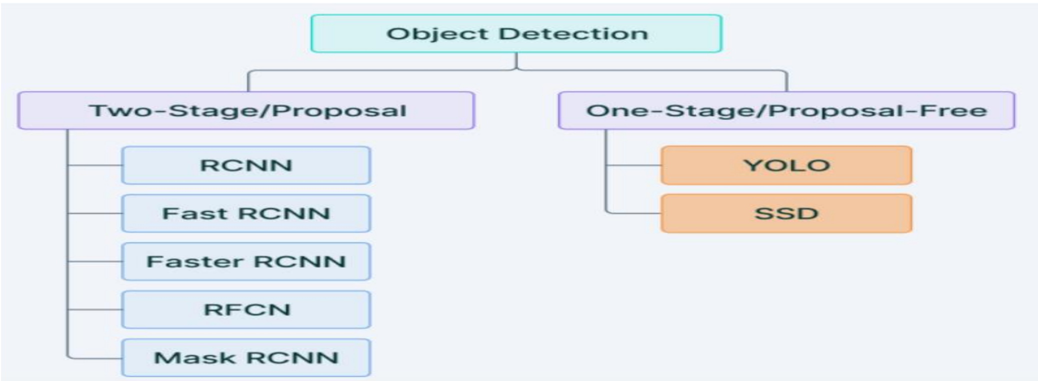

Figure 1. Flow chart of object detection

Since the popularization of deep learning in the early 2010s, there's been a continuous progression and improvement in the quality of algorithms used to solve object detection. We're going to explore the most popular algorithms while understanding their working theory, benefits, and their flaws in certain scenarios

R-CNN – Object Detection Algorithm

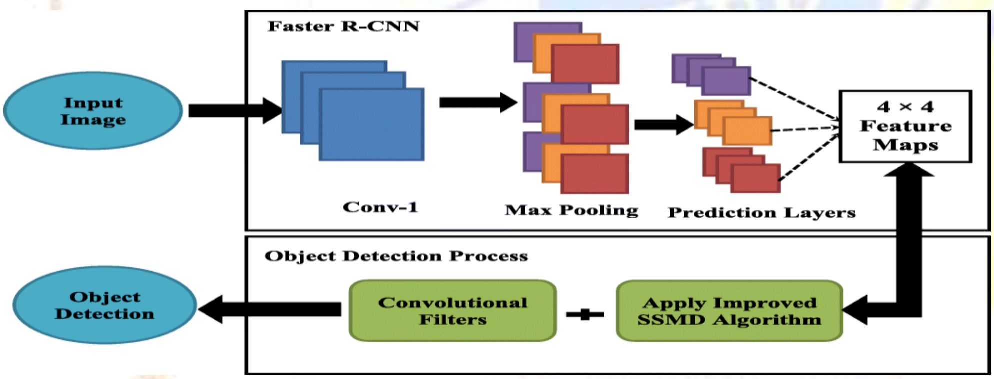

Figure 2. object detection using R-CNN

The working procedure of the selective search algorithm to select the most important regional proposals is to ensure that you generate multiple sub-segmentations on a particular image and select the candidate entries for your task. The greedy algorithm can then be made use of to combine the effective entries accordingly for a recurring process to combine the smaller segments into suitable larger segments Once the selective search algorithm is successfully completed, our next tasks are to extract the features and make the appropriate predictions. We can then make the final candidate proposals, and the convolutional neural networks can be used for creating an n-dimensional (either 2048 or 4096) feature vector as output. With the help of a pre-trained convolutional neural network, we can achieve the task of feature extraction with ease.

SSD – Object Detection Algorithm

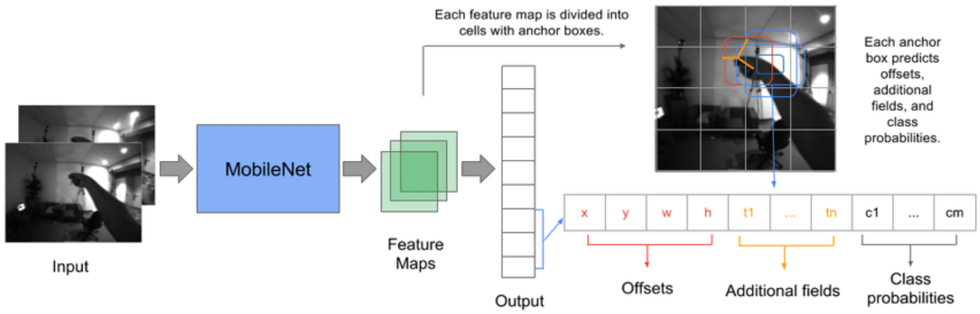

Figure 3. Object detection using SSD 1

The proposed system model. 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 model given in Fig.3

The single-shot multi-box detector architecture can be broken down into mainly three components. The first stage of the single-shot detector is the feature extraction step, where all the crucial feature maps are selected. This architectural region consists of only fully convolutional layers and no other layers. After extracting all the essential feature maps, the next step is the process of detecting heads. This step also consists of fully convolutional neural networks.

YOLO – Object Detection Algorithm

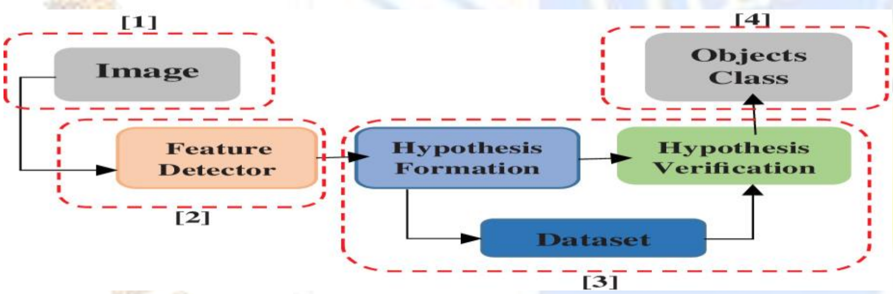

Figure 4. Object detection using YOLO

The YOLO architecture utilizes three primary terminologies to achieve its goal of object detection. Understanding these three techniques is quite significant to know why exactly this model performs so quickly and accurately in comparison to other object detection algorithms. The first concept in the YOLO model is residual blocks. In the first architectural design, they have used 7×7 residual blocks to create grids in the particular image.

Each of these grids acts as central points and a particular prediction for each of these grids is made accordingly. In the second technique, each of the central points for a particular prediction is considered for the creation of the bounding boxes. While the classification tasks work well for each grid, it's more complex to segregate the bounding boxes for each of the predictions that are made. The third and final technique is the use of the intersection of union (IOU) to calculate the best bounding boxes for the particular object detection task.

SSD vs YOLO

Unlike YOLO SSD does not divide the image into grids of random size. For every location of the feature map, it predicts the offset of predefined anchor boxes (default boxes). Relative to the corresponding cell, each box has a fixed size, proportion, and position. In a convolutional manner, all the anchor boxes cover the entire feature map. Anchors of SSD are slightly different from the anchors of YOLO. Because YOLO makes all the predictions from a single grid, and the size of anchors used by YOLO ranges from dimensions of one grid cell to the dimensions of the entire picture. The anchors of SSD specialize its detector for distinct feasible viewpoints and dimensional ratios of its target shapes, but not enough on the size of targets. The calculation for the anchors of SSD uses a simple formula, while the anchors of YOLO are calculated by applying k-means clustering on the training data. SSD doesn't use confidence score, but YOLO calculates it to show the faith in predicted results. A unique background class is employed by SSD for this work. A low value of confidence score in YOLO is equivalent to the predicted output of background class in SSD. Both indicate that for the detector, the possibility of getting a target is null.

Result

YOLO (You Only Look Once)

It only uses a single appearance at the image to detect several things. It is hence known as YOLO, or you only look once. The detecting speed can be determined after just one quick look at the image (45 fps). The FPS of Quick YOLOv1 is 125. This is typically another progressive deep learning object detection method that was published in 2016 CVPR and received close to 2000 citations.

Yolo creates a grid out of the image. Certain numbers, such as class probabilities and bounding box parameters, are computed for each grid. The model works by first dividing the input image into a grid of cells. If the centre of a bounding box falls within a cell, that cell is responsible for predicting a bounding box. A bounding box with the x, y, width, height, and confidence is predicted for each grid cell. Additionally, a class prediction is based on every cell.

The YOLO algorithm is among the best in detecting objects for the following reasons:

Speed: This method increases the speed of detection because it can anticipate things in real-time.

High accuracy: YOLO is a predictive method that gives accurate results with few background mistakes.

It has been put to use in numerous applications to find out about people, animals, parking meters, and traffic lights.

Figure 5. Result of YOLO methods 1

SSD (single-shot detector)

Features of SSD as follow:

Small convolutional filters to predict object classes and offsets to default boundary boxes.

Separate filters for default boxes to handle the difference in aspect ratios.

Multi-scale feature maps for object detection.

SSD can be trained end-to-end for better accuracy. SSD makes more predictions and has better coverage on location, scale, and aspect ratios. By removing the delegated region proposal and using lower resolution images, the model can run at real-time speed and still beats the accuracy of the state-of-the-art Faster R-CNN.

Figure 6. Result of SSD.

References

[1] 2. D. He, Z. Yao, Z. Jiang, Y. Chen, J. Deng and W. Xiang, "Detection of Foreign Matter on High-Speed Train Underbody Based on Deep Learning," *IEEE Access* , pp. 183838 - 183846, 2019.

[2] B. Zhang, C. Zhang, G. Li, L. Lin, C. Zhang, F. Wang and W. Ya, "Multispectral Heterogeneity Detection Based on Frame Accumulation and Deep Learning," *IEEE Access (Volume: 7),* pp. 29277 - 29284, 05 March 2019.

[3] I. O. D. U. W.-A. A. M.-L. Segmentation, "Improving Object Detection Using Weakly-Annotated Auxiliary Multi-Label Segmentation," *IEEE Access (Volume: 9),* pp. 161283 - 161291, 01 December 2021.

[4] I. S. M. A. L. K. CHUAN, "PAE: Portable Appearance Extension for Multiple Object Detection and Tracking in Traffic Scenes," *IEEE Access (Volume: 10),* pp. 37257 - 37268, 17 March 2022.

[5] E. Güney, C. Bayilmiş and B. Çakan, "An Implementation of Real-Time Traffic Signs and Road Objects Detection Based on Mobile GPU Platforms," *IEEE Access (Volume: 10),* pp. 86191 - 86203, 16 August 2022.

[6] G. Li, Z. Ji, X. Qu, R. Zhou and D. Cao, "Cross-Domain Object Detection for Autonomous Driving: A Stepwise Domain Adaptative YOLO Approach," *IEEE Transactions on Intelligent Vehicles,* pp. 603 - 615, 06 April 2022 .

[7] D. G. Lema, O. D. Pedrayes, R. Usamentiaga, P. Venegas and D. F. Garcí, "Automated Detection of Subsurface Defects Using Active Thermography and Deep Learning Object Detectors," *IEEE Transactions on Instrumentation and Measurement,* p. 4503213, 22 April 2022.

[8] S. Noor, M. Waqas, M. I. Saleem and H. N. Minhas, "Automatic Object Tracking and Segmentation Using Unsupervised SiamMask," *IEEE Access (Volume: 9),* pp. 106550 - 106559, 28 July 2021 .

[9] Y. Dai, W. Liu, H. Wang, W. Xie and K. Long, "YOLO-Former: Marrying YOLO and Transformer for Foreign Object Detection," *IEEE Transactions on Instrumentation and Measurement (Volume: 71),* p. 5026114, 15 November 2022.

[10] A. Koirala, M. Jha, S. Bodapati, A. Mishra, G. Chetty and P. Kishor, "Deep Learning for Real-Time Malaria Parasite Detection and Counting Using YOLO-mp," *IEEE Access (Volume: 10)*, pp. 102157 - 102172, 21 September 2022 .

[11] N. Zarei, P. Moallem and M. Shams, "Fast-Yolo-Rec: Incorporating Yolo-Base Detection and Recurrent-Base Prediction Networks for Fast Vehicle Detection in Consecutive Images," *IEEE Access (Volume: 10)*, pp. 120592 - 120605, 14 November 2022 .

[12] Z. Chen, K. Wu, Y. Li and M. Wang, "SSD-MSN: An Improved Multi-Scale Object Detection Network Based on SSD," *IEEE Access (Volume: 7)*, pp. 80622 - 80632, 14 June 2019.

[13] T. Ye, Z. Zhang, X. Zhang and F. Zhou, "Autonomous Railway Traffic Object Detection Using Feature-Enhanced Single-Shot Detector," *IEEE Access (Volume: 8)*, pp. 145182 - 145193, 10 August 2020.

[14] X. Lu, J. Ji, Z. Xing and Q. Miao, "Attention and Feature Fusion SSD for Remote Sensing Object Detection," *IEEE Transactions on Instrumentation and Measurement* , p. 5501309, 22 January 2021 .

[15] Z. Lyu, H. Jin, T. Zhen, F. Sun and H. Xu, "Small Object Recognition Algorithm of Grain Pests Based on SSD Feature Fusion," *IEEE Access (Volume: 9)*, pp. 43202 - 43213, 17 March 2021 .

[16] K. B. Lee and H. S. Shin, "An Application of a Deep Learning Algorithm for Automatic Detection of Unexpected Accidents Under Bad CCTV Monitoring Conditions in Tunnels," *2019 International Conference on Deep Learning and Machine Learning in Emerging Applications (Deep-ML)*, no. 21 October 2019, pp. 7281-7291, 26-28 August 2019.

[17] R. Kim, G. Kim, H. Kim, G. Yoon and H. Yoo, "A Method for Optimizing Deep Learning Object Detection in Edge Computing," *2020 International Conference on Information and Communication Technology Convergence (ICTC)*, no. 21 December 2020, 21-23 October 2020.

[18] 2. A. 3. T. 4. S. 5. D. 6. S. 7. P. 1Anushka, "Object Detection using Deep Learning: A Review," *Journal of Physics: Conference Series*, 2020.