International Journal of Computing and Information Technology Vol 1, No. 2, 2022
ISSN: 2790-170X DOI: 10.56028/ijcit.1.2.11

Overview of YOLO Object Detection Algorithm

Chengjuan Wan, Yuxuan Pang, Shanzhen Lan*

Communication University of China

awanchengjuan@cuc.edu.cn

Abstract

As an important research direction in the field of computer vision, object detection has developed rapidly and many kinds of mature algorithms emerged. The series of YOLO (You Only Look Once) algorithms implement one-stage detection based on regression ideas, which showing preeminent in speed and owning strong generalization on a variety of datasets. This paper will give a simple introduction to the current mainstream deep learning object detection algorithm, then focus on combing the principle and optimizational process of the series of YOLO algorithms, summarize the latest breakthroughs in YOLO algorithm, Hopefully that can provide reference for the research of related topics.

Keywords

object detection; deep learning; YOLO.

ISSN: 2790-170X DOI: 10.56028/ijcit.1.2.11

1. Object detection algorithm

Traditional object detection algorithms generate candidate regions by exhaustive sliding window and then feature extraction for machine learning classification with poor performance in speed and accuracy. With the rapid development of AI technology, it has been gradually replaced by deep learning methods.

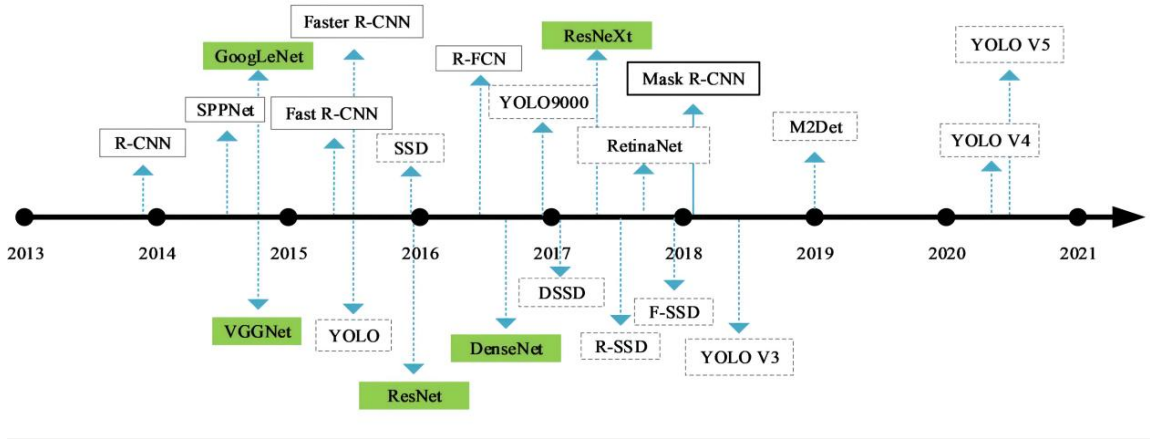

Fig.1. Timeline of object detection algorithm

1.1. Two-stage detection algorithm based on region ideas

The two-stage detection algorithm is represented by the series of R-CNN networks, since 2014 using region proposal and CNN instead of sliding window and manual design features to achieve RCNN framework construction, SPP-NET、Fast RCNN have appeared in succession. To make up for the time-consuming shortcoming that selectively search candidate boxes, Shaoqing Ren proposed the Faster RCNN algorithm which applies Regional Proposal Network to replace selective search to generate candidate regions and realizes weight sharing, ultimately forms a typical end-to-end two-stage detection process. Nowadays two-stage detection algorithm either includes FPN, Cascade-RCNN, Libra RCNN, Hybrid Task RCNN, PSS-Det and so on, all of which perform well in aspects of accuracy and recall.

1.2. One-stage detection algorithm based on regression ideas

The one-stage detection algorithm is represented by YOLO, SSD, RetinaNet. In 2016 Redmon proposed YOLOv1, which produces candidate boxes for classification and bounding box regression without the output process of intermediate regions, bringing about real-time improvement at the sacrifice of a small amount of precision rate. In the same year, Liu proposed the SSD algorithm that combines YOLO with Faster RCNN to solve the localization accuracy problem, moreover adding multi-scale feature maps for prediction. Since then, the series of YOLO algorithms continue to develop, will be highlighted thereinafter .

1.3. Anchor-free object detection algorithm

The above mainstream algorithms actualize object detection based on bounding boxes, which

ISSN: 2790-170X DOI: 10.56028/ijcit.1.2.11
include many problems: the irregular object appearance makes the bounding box contain some non-object areas, interfering to detection result; the setting of hyper- parameters such as the number， size and width ratio of bounding box is needed to adjust according to the datasets; the large number of bounding boxes may lead to the imbalance of positive and negative samples, affecting the training effect. In 2018, Law H proposed CornerNet, converting the object detection problem into a key-point detection problem, which uses a single convolutional neural network to predict two key points in the upper left and lower right corners of the object, thus obtaining the prediction box. CornerNet focused on object edges, in 2019 Duan K proposed CenterNet, increased detection of the center site and then regressed to other properties of the object through the location of the center site. Up to this day, the CornerNet-Squeeze, FCOS, and TTFNet algorithm have all performed well.

2. YOLO algorithm

In order to make up for the shortcomings of Faster RCNN be difficult to meet real-time requirements in detection speed, Joseph Redmon, Santosh Divvala, Ali Farhad and others proposed YOLO, which regards the object detection framework as a spatial regression problem. A single neural network can get bounding box and class prediction from the complete image after one operation. YOLO experienced the version update from v1 to v5 , has become one of the mainstream framework of object detection.

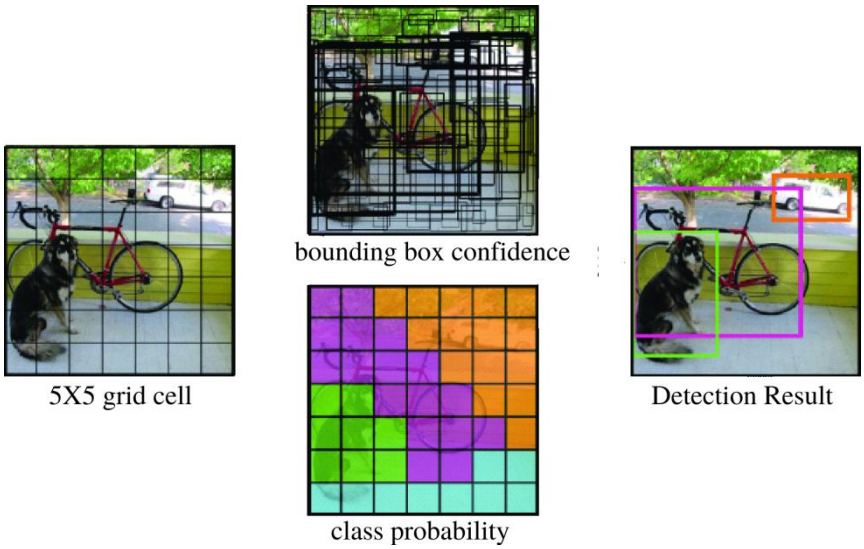

Fig.2. YOLO algorithm implementation process

2.1. YOLOv1

YOLOv1 achieves the prediction of object classes and bounding box regression on a complete image directly through the CNN. Its network structure is built based on the GoogLeNet model. Firstly, the input image is fixed to a uniform size (448x448), the input image is divided into SxS grid cells, each of them is responsible for detecting the object falling on it and predicting the confidence, class and location of the object. Secondly, extracting the feature in input images using CNN. And finally, the optimal result is obtained by processing Non-Maximum

ISSN: 2790-170X DOI: 10.56028/ijcit.1.2.11
Suppression (NMS). Each grid cell divided by YOLO detects an object and transforms the detection bounding box by regression so that the framework can extract features directly from the input image to predict the object bounding boxes and class probabilities.

The YOLO detection system is designed to divide the image of the input 448x448x3 into grid of 7x7, and whose calculation of the predicted output tensor is expressed as:

$S \times S (B \times 5 + C)$

Where SxS represents the number of grids divided by the input image, corresponding to the feature graph resolution; B represents the number of bounding boxes generated per grid; 5 represents the number of predicted parameters (x, y, w, h, confidence), and C represents the identified species that can be detected (20). Compared to the Fast RCNN of 0.5 fps and Faster RCNN of 7 fps, YOLO operates at 45 fps, with a large improvement in running speed, but still some deficiencies in prediction accuracy, prone to more positioning errors especially for small objects.

2.2. YOLO v2

YOLOv2, relative to the v1 version, mainly improves on the prediction accuracy, speed, and the number of identified objects. The objects detected by YOLOv2 extend to 9 000 species. YOLOv2 optimizes in the following aspects： It uses a simpler feature extraction network DarkNet19 to replace the GoogLeNet network; Batch normalization (BN) was introduced to strengthen the convergence rate of the network, and enhanced generalization; Trains a high-resolution classifier to accommodate higher-resolution images; Removes a pooling layer to increase the output resolution of the convolutional layer; The ImageNet classification dataset and the COCO detection dataset were jointly trained using WordTree; A pass-through layer was added, by connecting the last output 3x3x512 layer with the preceding convolutional layer to contact high resolution image features and low resolution image features, features were acquired from the earlier layers at a resolution of 26x26; To adapt to more pictures of different sizes, after every ten iterations, select a new resolution (from 320x320 to 608x608) for the operation; The full connection layer was removed and the prior box (or anchor boxes) was automatically searched for using the k-means clustering algorithm, improve the detection performance.

2.3. YOLOv 3

In 2018, the original author proposed the YOLOv3 algorithm, which inherits the ideas of YOLOv1 and YOLO9000 and achieves a balance of speed and detection accuracy. In terms of network structure, YOLOv2 canceled all the fully connected layers of the first generation YOLOv1, while YOLOv3 further abolished all the most common pooling layers in the convolutional neural network, and the original pooling layer used to reduce the feature size was changed by increasing the step length of the original convolutional core, which greatly improves the speed. Another important improvement of YOLOv3 is the ability to output three different size feature maps of both 13x13, 26x26, and 52x52, enhancing the detection of small objects, but also weakening large objects. Finally, YOLOv3 no longer uses Softmax to classify each box, and instead uses multiple separate logistic classifiers, operating only the one of the obtained anchor frame with the highest object likelihood score.

ISSN: 2790-170X DOI: 10.56028/ijcit.1.2.11

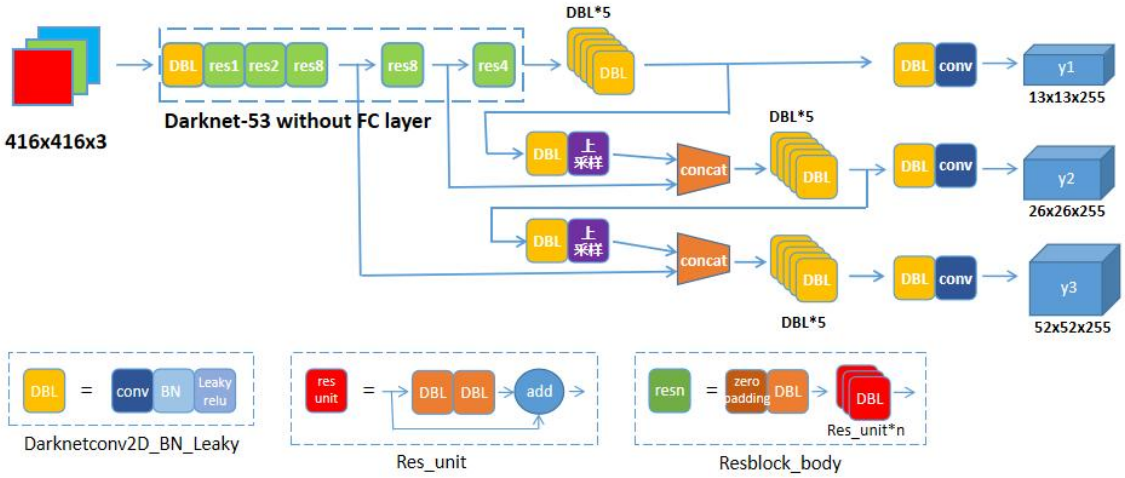

Fig.3. YOLOV3 structure

2.4. YOLOv4

In April 2020, Alexey improved and proposed a new algorithm for efficient object detection based on YOLOv3, YOLOv4. It is characterized by integration, including new data enhancement methods Mosaic and self-Adversarial Training (SAT) method, proposed improved SAM and PAN, and cross-small batch normalization (Cross mini-Batch Normalization, cmBN). Including Head, Neck for SPP, and PAN from YOLOv3, and Backbone from CSPDarkNet53.

The YOLOv4 is split into 4 sections including input terminal, BackBone, Neck, prediction part.Among them, the input side mainly includes Mosaic, cmBN, SAT; the backbone network includes the CSPDarknet53 network, Mish activation function, Dropblo-ck; the Neck part includes the SPP module, FPN + PAN structure; the prediction part is mainly the improved loss function CIOU_Loss, and the bounding box filtered nms becomes DIOU_nms.

2.5. YOLOv5

After two months after YOLOv4, some researchers launched the YOLOv5 algorithm. In terms of accuracy metrics, its performance is equal to YOLOv4, far exceeds v4 in speed, and the model size(27MB) is also very small than YOLOv4(245MB) . It has a strong advantage in model deployment. The YOLOv5 structure is slightly similar to the YOLOv4, but it's different. Its input adopts Mosaic data enhancement, adaptive anchor frame calculation and adaptive image scaling; Framework includes Focus structure, Backbone of CSP structure and Neck of FPN structure. Through the improvement of YOLO series algorithm and comparison with RCNN series algorithm, YOLO achieves "you only look once" relative to the extraction and classification of RCNN series; YOLO unifies the detection as a regression problem, while RCNN divides the detection results into object class (classification problem) and object location (regression problem). A range of improved approaches give YOLO a lead in the speed of object detection.

ISSN: 2790-170X DOI: 10.56028/ijcit.1.2.11

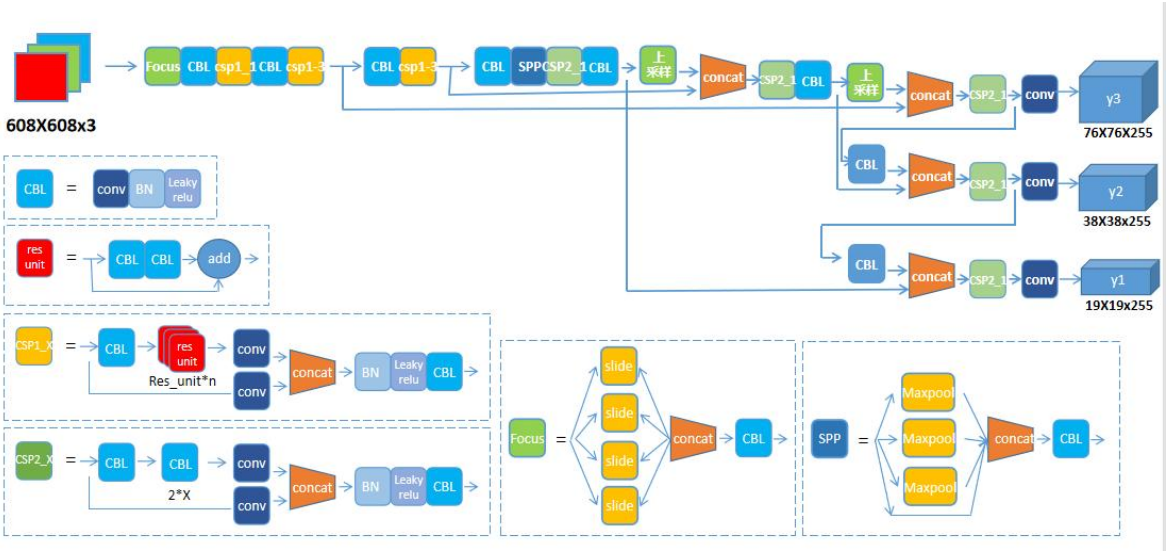

Fig.4. YOLOV5s structure

3. Summary and Outlook

This paper focuses on the field of object detection and introduces the development and optimization process of the series of YOLO algorithms. In order to meet its application optimization in light weight, small object detection, precision speed and so on, extension algorithms such as PP-YOLOv2, PP-YOLOv2, P P-YOLO Tiny, YOLO-Fastest and YOLObile have also emerged continuously, worthing further research.

Reference

[1] ZHOU X Y, GONG W, FU W L, et al. Application of deep learning in object detection[J]. In Proceedings of the IEEE/ACIS 16th International Conference on Computer and Information Science, 2017, 132(5) : 631-634.

[2] QIAN X, LIN S, CHENG G, et al. Object detection in re-mote sensing images based on improved bounding box re-gression and multi-level features fusion[J]. Remote Sens, 2020, 12(1) : 143-164.

[3] RUSSAKOVSK O, DENG J, SUH, et al. Imagenet large scale visual recognition challenge [J]. International Journal of Computer Vision, 2015, 115(3) : 211-252.

[4] LIU L, OUYANG W, WANG X, et al. Deep learning for generic object detection: a survey [J]. International Journal of Computer Vision, 2020, 128: 261-318.

[5] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and se-mantic segmentation[J]. IEEE Conference on Computer Vision and Pattern Recognition, 2014, 81(1) : 580-587. .

[6] UIJLINGS J, SANDE K, GERS T, et al. Selective search for object recognition[J]. International Journal of Computer Vision, 2013, 104(2) : 154-171.

[7] COETES C, VAPNIK V. Integrated series in information systems[M]. Berlin: Springer, 1995: 207-235.

ISSN: 2790-170X DOI: 10.56028/ijcit.1.2.11
[8] REN S, HEK, GIRSHICK R, et al. Faster RCNN: towards real-time object detection with region proposal networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39: 1137-1149.