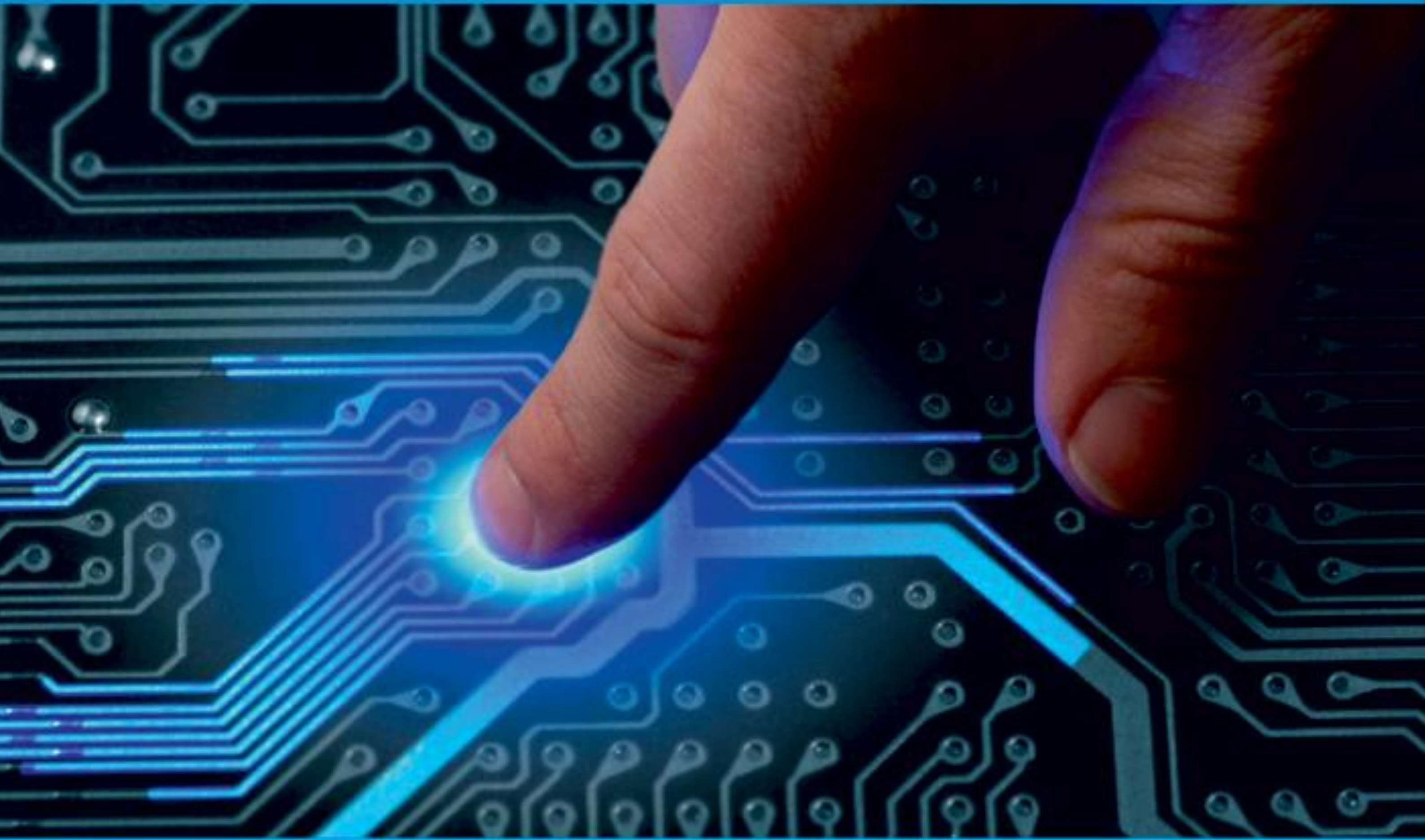
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 10, Issue 5, May 2022

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.165

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 5, May 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1005224 |

Object Detection and Labeling

Pavanlingraj, Prof. Md. Irshad Hussain B

Student, Master of Computer Applications, University B.D.T College of Engineering, Davanagere, Karnataka, India

Assistant Professor, Department of Master Applications, University B.D.T College of Engineering, Davanagere, Karnataka, India

ABSTRACT: Object detection is thought to be one of the most difficult tasks in computer vision. Object detection has been a major area of research for the past 20 years. This is because it is one of the most important parts of computer vision. An object detection system finds real-world objects in a digital image or video, such as people, animals, cars, buses, etc. To find an object in an image or video, the system needs a model database and a feature detector. This paper gives an overview of the different methods or techniques that can be used to find an object in an image, figure out where it is, classify it, pull out its features, information about how it looks, and much more. Our review starts with a short history of deep learning, which has been used by many people. Its goal is to quickly and accurately find and identify a large number of objects in a given image that belong to predefined categories.

I.INTRODUCTION

People are very good at finding and pointing out objects in an image. The human visual system is fast and accurate, and it can do complicated things like recognise many objects and spot obstacles with little conscious thought. With large amounts of data, faster GPUs, and better algorithms, it is now possible to train computers to find and classify multiple objects in an image with high accuracy[1].

It tries to find the object of interest in an image, narrow down the category, and give the bounding box for each object. It is a requirement for more advanced computer vision tasks like following a target, recognising events, and analysing behaviour. It has helped with automated driving, retrieving videos and images, smart video surveillance, industrial inspection, and other things[2].

Traditional feature extraction methods have six steps: preprocessing, window sliding, feature extraction, feature selection, feature classification, and postprocessing. These methods can be used for a wide range of tasks. Some of its flaws include: small data size, low portability, lack of relevance, high time complexity, window redundancy, lack of robustness for changes in diversity, and only acceptable performance in a few simple situations. Krizhevsky and his team came up with the convolutional neural networks-based AlexNet image categorization model (CNN)[3]

II.LITERATURE SURVEY

Sutskever [4] wrote about this in his Survey and Performance Analysis of Deep Learning-Based Object Detection in Challenging Environments. On three sets of difficult datasets, they tested how well the best object detection methods work right now. The shcemis will study how well-trained object detection algorithms work when they are put to the test. They tested how well they did on the datasets ExDARK, CURE-TSD, and RESIDE by using Faster R-CNN, Mask R-CNN, YOLO V3, Retina-Net, and Cascade Mask R-CNN.

Russakoysky [5], Target identification is one of the most important jobs in computer vision, and it has been a popular area of study and work for the past 20 years. Its goal is to quickly and accurately find and identify a large number of things in an image that belong to categories that have already been set. Based on how the models are trained, the algorithms can be put into two groups: There are two kinds of detection algorithms: ones with one step and ones with two. This publication goes into detail about the algorithms that are used at each level. Then, different sample algorithms and public and private datasets that are often used in target detection are compared and contrasted. Lastly, problems with finding targets are talked about.

Girshik[6], Scene analysis, video surveillance, robotics, and self-driving cars are just some of the many uses that have led to a lot of research in the field of computer vision in the last ten years. Visual recognition systems, which include things like classifying, localising, and detecting pictures, are at the heart of all of these applications and have gotten a lot of research attention. These algorithms for visual identification are very good because neural networks, especially deep learning, have come a long way in recent years. One area where computer vision has been very successful is object recognition.

K.M.Zjang [7] The version of the best object detection algorithms that can be used on hard datasets is shown. Since finding objects in harsh environments is similar to finding objects in general, they will use the same evaluation criteria to describe the results. It's clear that there's a lot of room for improvement in all of the datasets they used. With an AP of 0.67, YOLO V3 did the best job with the ExDark dataset. Cascade Mask R-CNN gives the best score for CURE-TSD.
Girshik's [8] survey covers most object detection applications for maritime surveillance and self-driving ships. In the past few years, a lot of deep learning-based models for finding objects in the ocean have been proposed. However, there aren't any common evaluation criteria, so it's hard to compare the different improved models. This study summed up the benefits of the computer vision milestone model and showed how the single-stage model and the multistage model could be used in different ways based on how the marine environment works.

Girshik ,r., Sun
[9] Deep learning-based object detection has become a popular area of research in recent years because it can learn so much and is good at dealing with occlusion, scale changes, and changes in the background. This study gives an in-depth look at deep learning-based object detection frameworks that deal with different sub-problems, such as occlusion, clutter, and low resolution, by modifying R-CNN in different ways. The talk starts with generic object detection pipelines, which are the basis for other activities in the same area. Then, three other common tasks are briefly talked about: finding things that stand out, finding faces, and finding people. Lastly, to get a full picture of the object detection landscape, you should suggest a number of possible next steps.

Christopher Sager etc., [10] AI and machine learning have come a long way, which has made it much easier to build complex CV systems. A big problem is that these systems still need to be trained with a lot of annotated examples and under supervision. ILS's most important job here is to add high-quality data that has been checked by humans to the knowledge base. ILS are the main link between smart machines and human experts. They let domain-specific knowledge be put into the learning bases of CV systems.

III.METHODOLOGIES

It can be roughly put into two groups: the single-stage detection algorithm based on region proposal and the two-stage detection algorithm based on regression.

3.1 Two-Stage Target DetectionFramework

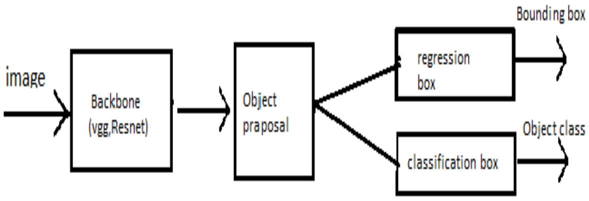


Fig:-exhibits the basic architecture of two-stage detectors[11]

3.2 TWO-STAGE TARGET DETECTION

1.R-CNN[11]
Girshick came up with the R-CNN algorithm in 2014. It is the first real model for detecting targets that is based on convolutional neural networks. Rich feature hierarchies for accurate object detection and semantic segmentation" describes the "R-CNN" or "Regions with CNN," which was one of the first breakthroughs in using CNNs in an object detection system. The model starts by using the Selective Search to pull out about 2000 proposed regions from each image to be detected. SVMs classification takes a long time for R-CNN. Fast RCNN takes features from the whole input image and

then sends them to the region of interest (ROI) pooling layer to get fixed-size features that are used as inputs for the classification and bounding box regression layers that come next.

The features are taken from the whole image and sent to CNN at the same time so that it can classify and locate the image. Compared to RCNN, which sends each region proposal to CNN, Fast RCNN saves a lot of time for CNN to process and a lot of disc space to store a lot of features. Making suggestions for the region. The R-CNN uses selective search to come up with about 2,000 ideas for each image. In object detection, the selective search method uses basic bottom-up grouping and salience cues to quickly give more accurate candidate boxes of any size and to limit the search space.

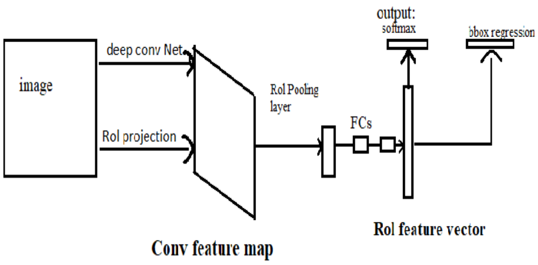


Fig:R-CNN architecture[11]

CNN is used for deep feature extraction. At this stage, each region suggestion is stretched or cut to a fixed resolution. The CNN module is then used to pull out a 4096-dimensional feature as the final representation. Because CNNs have a strong ability to express themselves and a hierarchical structure, each area suggestion can have a high-level, semantic, and strong representation of its features. Localization and classification It is important to improve the quality of candidate bounding boxes and use a deep architecture to get high-level features. Bishop et al. [12] come up with a unique detector called RefineDet. It takes the best parts of one-stage and two-stage detectors and fixes their flaws. It is more accurate than two-stage detectors while keeping the efficiency of one-stage detectors.

2.Spatial Pyramid Pooling (SPP) [13]

The model proposed in 2015 was called Spatial Pyramid Pooling (SPP). The problem with R-low CNN's ability to find things and the need for image blocks with a fixed size has now been solved. After the original image has been through the convolution layer, this method takes the features from the areas defined on the feature map and does only one convolution calculation. After the last convolutional layer, the spatial pyramid pooling layer is added, and the region proposal feature is sent through it to get the fixed-size feature vector. Spp-Net, unlike R-CNN, only extracts features from the whole image once, so it doesn't have to do multiple calculations. It has the same problems as R-CNN, though: 1) Multi-step training procedures are difficult. Separate SVM classifiers and extra regressors must be trained.

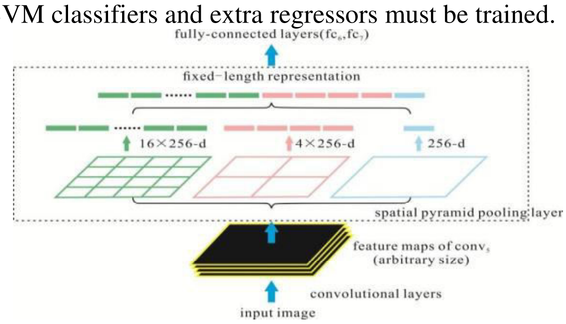

fully-connected layers(fc_6,fc_7)
fixed-length representation
16×256-d
4×256-d
256-d
spatial pyramid pooling layer
feature maps of $conv_5$
(arbitrary size)
convolutional layers
input image

Fig:SPP-Net architecture [13]

SPP (Spatial Pyramid Pooling)[14] is a pooling layer that gets rid of the network's fixed-size restriction. This means that a CNN can work without an input image of a fixed size. They put an SPP layer on top of the last convolutional layer. The SPP layer collects the features and makes outputs of a fixed length, which are then sent to the fully-connected layers (or other classifiers). To put it another way, they gather information at a higher level of the network structure (between the convolutional and fully connected layers) so that they don't have to crop or warp the image at the beginning. The SPP-power net can also be used to find things. SPP-net is used to make a single set of feature maps for the whole image. Then, features from random sections (sub-images) are combined to make fixed-length representations for training the detectors.

3.Fast R-CNN[15]
Fast R-CNN is a new model for identifying objects that is better in a number of ways than R-CNN. Instead of extracting CNN features separately for each area of interest, Fast R-CNN combines them into a single forward pass over the image. Both the forward and backward passes share the same memory and processing power.

The Faster R-CNN model was made by a group of Microsoft researchers. Faster RCNN is a deep convolutional network that looks to the user like a single, unified network from beginning to end. It is used to find objects. The network can quickly and accurately guess where different things are. Fast-RCNN, unlike RCNN and SPPNet, lets you learn both classification and regression from start to finish. Between the last conv layer and the fully connected layer, Fast-RCNN also used ROI pooling to wrap feature maps of any size. So, Fast-RCNN took three times as long as other networks to train. Even after, the speed of Fast-RCNN goes up. Fast-RCNN uses ways to suggest regions from the outside. In the Fast-RCNN architecture, it turns out that the problem is the computing. Study after study has shown that CNNs are very good at finding objects in Convolution layers (Zhou et al., 2015; Cinbis et al., 2017; Oquab et al., 2015). In a later, fully connected layer, this is turned off.

R-CNN takes a long time to classify SVMs because for each area suggestion, it does a ConvNet forward pass without sharing computation. Fast RCNN takes features from the whole image and sends them through the region of interest (ROI) pooling layer to the fully connected classification and bounding box regression layers. The features are taken from the whole image and sent to CNN at the same time so that it can classify and locate the image. Compared to R-CNN, which sends each region proposal to CNN, Fast R-CNN saves a lot of time for CNN processing and allows for a large number of features to be stored on a big disc[15].

3.3 One-STAGE TARGET DETECTION FRAMEWORK

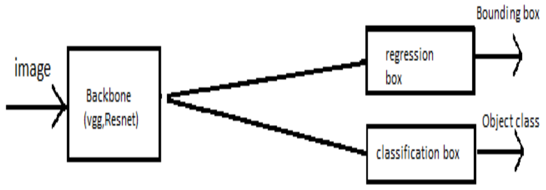


Fig:shows the basic architecture of one-stage detectors.[11]

3.4One-Stage Target Detection Algorithm

1. YOLOv1(You Only LookOnce,v1)2016 [16]
In 2016, Joseph Redmon came up with the YOLOv1 method for finding objects. The YOLOv1 detection model doesn't need the method for getting region proposals. The whole system for figuring out what is going on is just a simple CNN network. The basic idea is to give the network the full graph as input and have the output layer report the location and type of the bounding box. First, an image is broken up into a S*S grid, with each grid cell predicting B bounding boxes and their confidence scores. That is, each cell predicts B*(4+1) values in total. The tests showed that YOLO had trouble with localization. In fact, localization errors made up most of the prediction errors.

Fast R-CNN makes a lot of false positives in the background, while YOLO makes three times YOLOv1 is a model for finding things that only has one step. Object detection is modelled as a regression problem with bounding boxes and class probabilities that are in different places. It also predicts all of an image's bounding boxes at once for all classes. This means that the network takes into account the whole picture and all of its parts. "You Only Look Once," or "YOLO," was a big step forward in the field of finding things. It was the first time object detection was looked at as a regression problem.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 5, May 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1005224 |

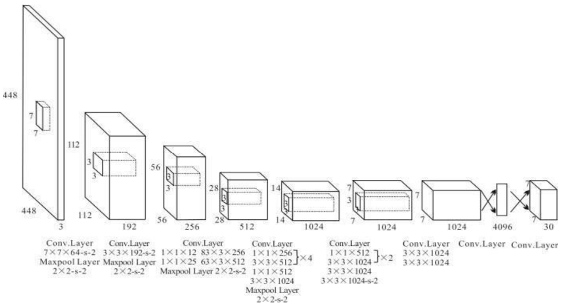


Fig:YOLOv1 architecture[17]

With this model, you only need to look at an image once to guess what things are there and where they are. Unlike the two-stage detector method, YOLO uses a single neural network to predict class probabilities and bounding box coordinates from an entire image in a single run. Because the detection pipeline is really just one network—think of it as an image classification network—it can be optimised from start to finish. YOLO learns representations of objects that can be used in many different situations. When trained on natural photos and tested on art, YOLO does a lot better than leading detection algorithms like DPM and R-CNN.
YOLO is called "you only look once" because its prediction uses 11 convolutions. The size of the prediction map is the same as the size of the feature map that came before it.

2. YOLOv2(You Only Look Once,v2)[18]
In December 2016, Joseph Redmon and Ali Farhadi released the second version of their object detection model, YOLO (You Only Look Once). It's a one-step design that goes from picture pixels to bounding box coordinates and class probabilities in a single step. The old YOLO architecture has many problems when compared to modern methods like Fast R-CNN. It has a low recall and a lot of trouble being used in the right place. So, the goal of this study is not only to fix YOLO's flaws but also to keep the speed of the architecture. Batch Normalization, high resolution classifier, Use Anchor Boxes For Bounding Boxes, Dimensionality clusters, Direct Location Problem, and other small changes are made to basic YOLO.

3.YOLOv3(You Only Look Once, V3)[19]
You Only Look Once, Version 3 (YOLOv3) is an object detection system that can find specific things in photos, videos, or live feeds. YOLOv3 is like YOLOv2 and YOLO, but it is more advanced. To make YOLO work, either the Keras or OpenCV deep learning libraries are used. A deep convolutional neural network teaches YOLO the features of an item so that it can find it. The third version of YOLO was made with the help of Joseph Redmon and Ali Farhadi. Artificial Intelligence (AI) algorithms use systems for classifying objects to figure out which objects in a class are interesting. Things in photographs are put into groups with things that have similar qualities. Other things are ignored unless told to do so.

4.SSD(Single Shot Detector)[20]
When multibox is used, a one-shot detector like YOLO only needs one shot to find many things in an image. SSD only needs one shot to recognise many objects in an image, while RPN-based techniques like the R-CNN series need two shots, one to suggest regions and the other to find the object in each suggestion. Because of this, SSD is much faster than two-shot RPN-based methods.

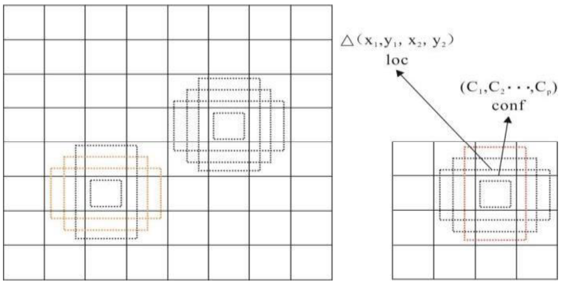


Fig : SSD architecture [20]

It has a system that can find objects much faster and more accurately. A quick look at how fast and accurately different object identification models work.

High speed and accuracy of SSD using relatively low resolution images is attributed due to following reasons

- Eliminates bounding box proposals like the ones used in RCNN's
- Includes a progressively decreasing convolutional filter for predicting object categories and offsets in bounding box locations.

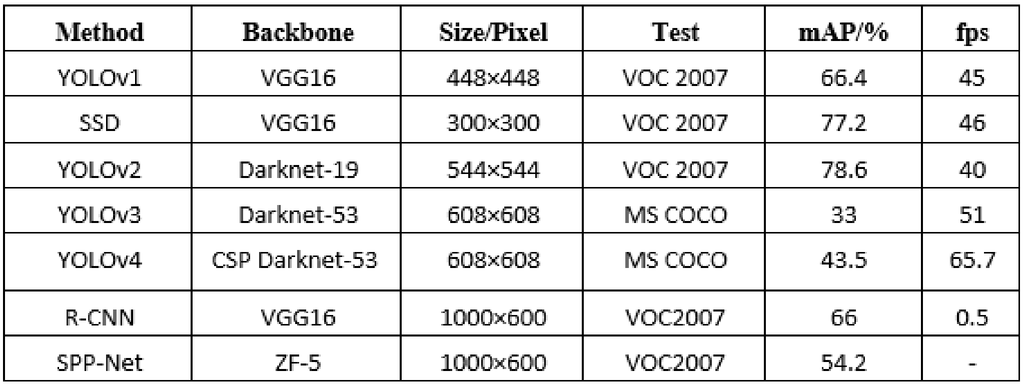
5.YOLOv4 [21]

Alexey Bochkovskiy, Chien-Yao Wang, and Hong-Yuan Mark Liao worked together to make the game YOLO v4. It came out in April 2020, and its improved AP and FPS make it stand out. YOLOv4 puts all of the training and real-time object detection on a single CPU. As an object detector with only one stage, YOLOv4 is more accurate and faster than R-CNN and Fast R-CNN.

The best thing about using YOLO is that it is so fast. It can process 45 frames per second. YOLO also knows how to represent objects in a more general way. This is one of the best algorithms for finding objects. It works about as well as the R-CNN algorithms.

There are a lot of deep learning frameworks on the market right now. Some important frameworks for deep learning have been talked about. The frameworks are compared based on their features, interface, support for deep learning models like convolutional neural networks, recurrent neural networks (RNN), Restricted Botltzmann Machine (RBM), and Deep Belief Network (DBN), support for multi-node parallel execution, the framework's developer, and licence.

TABLE .RESULT AND COMPARISON OF OBJECT DETECTION ALGORITHMS



Method	Backbone	Size/Pixel	Test	mAP/%	fps
YOLOv1	VGG16	448×448	VOC 2007	66.4	45
SSD	VGG16	300×300	VOC 2007	77.2	46
YOLOv2	Darknet-19	544×544	VOC 2007	78.6	40
YOLOv3	Darknet-53	608×608	MS COCO	33	51
YOLOv4	CSP Darknet-53	608×608	MS COCO	43.5	65.7
R-CNN	VGG16	1000×600	VOC2007	66	0.5
SPP-Net	ZF-5	1000×600	VOC2007	54.2	-

IV.CONCLUSION

Deep learning-based object detection has become a popular area of study in recent years because it can learn quickly and is good at dealing with occlusion, scale changes, and background changes.

It has been found that machine learning in object detection is a good way to deal with things like occlusion, location, scale transformation, and lighting. The machine learning method has done very well at many vision tasks, such as putting pictures into groups, identifying objects, and putting them into groups. In particular, the machine learning technique improves performance by telling apart sub-level features based on how they are classified in an image. The object detection system can tell if an object is there or not in certain situations and from certain camera angles. The many areas of object detection are put into specific and general groups with different goals in mind. Using different models, either explicitly or intuitively, to find objects. The many methods and strategies used by object detection systems.

REFERENCES

[1] Wu, R.B. Research on Application of Intelligent Video Surveillance and Face Recognition Technology in Prison Security. China Security Technology and Application. 2019,6: 16-19.

[2] Tian, J.X., Liu, G.C., Gu, S.S., Ju, Z.J., Liu, J.G., Gu, D.D. Research and Challenge of Deep Learning Methods for Medical Image Analysis. Acta Automatica Sinica,2018, 44: 401-424.

[3] Jiang, S.Z., Bai, X. Research status and development trend of industrial robot target recognition and intelligent detection technology. Guangxi Journal of Light Industry, 2020, 36: 65-66.

[4] Krizhevsky, A., Sutskever, I., Hinton, G. ImageNet Classification with Deep Convolutional Neural Networks. Advances in Neural Information Processing Systems,2012, 25: 1097-1105.
[5] Russakovsky, O., Deng, J., Su, H., et al. ImageNet Large Scale Visual Recognition Challenge.International Journal of Computer Vision,2015, 115: 211-252.
[6] Girshick, R., Donahue, J., Darrel, T.,Malik, J. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. In: Computer Vision and Pattern Recognition.
Columbus.2014, pp. 580-587.
[7] He, K.M., Zhang, X.Y., Ren, S.Q., Sun, J. Spatial Pyramid Pooling in Deep convolutional Networks for Visual Recognition. IEEE Transactions on Pattern Analysis & Machine Intelligence,2015, 37: 1904-1916.
[8] Girshick, R. Fast R-CNN.In: Proceedings of the IEEE international conference on computer vision. Santiago.2015, pp. 1440-1448.
[9] Ren, S.Q., He, K.M., Girshick, R., Sun, J. Faster R-CNN: towards real-time object detection with region proposal networks. In: Advances in neural information processing systems.Montreal.2016, pp. 91-99.
[10] Redmon, J., Divvala, S., Grishick, R., Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In: Computer Vision and Pattern Recognition. Las Vegas.2016, pp. 779-788.computer vision applications. In Construction Research Congress 2016 ,vol 2, no.16, pp. 2573-2582.2016
[11]. Dietterich, T. G. Ensemble methods in machine learning. In International workshop on multiple classifiersystems Springer, Berlin, Heidelberg, vol 2, no.3, pp. 1-15.2000.
[12]. Bishop, C. M. . Pattern recognition and machine learning. Springer, vol 2, no.8, pp:23-29.2006
[13]. Han, C., Liu, X., Sinn, L. T., & Wong, T. T. .TransHist: Occlusion-robust shape detection in clutteredimages. Computational Visual Media, vol4,no. 2,pp. 161-172.2018
[14]. Li, C., Zhang, Y., &Qu, Y. Object detection based on deep learning of small samples. In AdvancedComputational Intelligence (ICACI), 2018 Tenth International Conference on .IEEE,vol2, no.3, pp. 449-454.2018, March.
[15]. He, K., Zhang, X., Ren, S., & Sun, J. . Deep residual learning for image recognition. In Proceedings of the IEEEconference on computer vision and pattern recognition vol 2, no.3, pp. 770-778.2016.
[16]. Erhan, D., Szegedy, C., Toshev, A., &Anguelov, D. Scalable object detection using deep neural networks.In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition vol 3, no 4,pp. 2147-2154.2014
[17]. Kamate, S., &Yilmazer, N. Application of object detection and tracking techniques for unmanned aerialvehicles. Procedia Computer Science, vol61,no.3, pp. 436-441.2015
[18]. Kurian, M. Z., & MV, C. M. Various Object Recognition Techniques for Computer Vision. Journal of Analysis andComputation, vol7, no 1,pp. 39-47.2011.
[19] P. F. Felzenszwalb, R. B. Girshick, D. McAllester and D. Ramanan, “Object detection with discriminativeltrained part-based models,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, no. 9,pp. 1627–1645, 2010.
[20] C. Desai, D. Ramanan and C. C. Fowlkes, “Discriminative models for multi-class object layout,” InternationalJournal of Computer Vision, vol. 95, no. 1, pp. 1–12, 2011.
[21] L. Zhang, L. Lin, X. Liang and K. He, “Is YOLO doing well for pedestrian detection,” in Proc. of theEuropean Conf. on Computer Vision, Amsterdam, Netherlands, pp. 443–457, 2016.

INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor: 8.165

doi crossref

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

Scan to save the contact details

www.ijircce.com