OBJECT DETECTION TECHNIQUES BASED ON DEEP LEARNING: A REVIEW

Utkarsh Namdev, Shikha Agrawal and Rajeev Pandey

Department of Computer Science and Engineering, Rajiv Gandhi Proudyogiki Vishwavidyalaya Bhopal, 462033, M.P, India

ABSTRACT

Object detection is a computer technique that searches digital images and videos for occurrences of meaningful subjects in particular categories (such as people, buildings, and automobiles). It is related to computer vision and image processing. Two well-studied aspects of identification are facial and pedestrian detection. Object detection is useful in a wide range of visual recognition tasks, including image retrieval and video monitoring. The object detection algorithm has been improved many times to improve the performance in terms of speed and accuracy. "Due to the tireless efforts of many researchers, deep learning algorithms are rapidly improving their object detection performance. Pedestrian detection, medical imaging, robotics, self-driving cars, face recognition and other popular applications have reduced labor in many areas." It is used in a wide variety of industries, with applications range from individual safeguarding to business productivity. It is a fundamental component of driver assist systems and driverless cars, which allows automobiles to identify driving lanes and pedestrians to avoid any accidents.

KEYWORDS

Deep Learning, Convolutional Neural Networks, RCNN, YOLO.

1. INTRODUCTION

Researches were conducted in recent years to produce a realistic technique to speed up the development of profound learning methods. There's been several good advancements and continued revision of deep learning schemes. Object positioning is to identify all the visual effects in the image and combine the exact location of these visual effects. By using deep learning technology to recognize and locate objects, computer vision has reached a new high. Due to the significant instability of point of view, posture, size and lighting position, it is challenging to identify objects perfectly. Object detection plays a central role in the discovery and recognition of devices in scenes or images, e.g. To determine the number of cars on the road at a given time or the number of passengers waiting at the bus stop during peak times. Scene recognition plays an important role in the classification of areas affected by floods and the classification of land cover in specific locations. Among the specific concepts and measures thathave contributed to the rapid development of object search systems, the development of deep convolutional neural networks and GPU computing power should be attributed. Deep learning models are now widely used throughout the computer view field, including object recognition ingeneral and specific fields.

Image segmentation creates a pixel-level understanding of scene components, while image recognition simply emits a class label for each recognized object. The difference between object registration and these other tasks is its ability to locate elements in images or videos. As a result, we can count and track these things. Object registration in computer vision refers to the process of locating and registering objects in images or videos. The required main phases are function extraction, function processing and classification. Function extraction is an important part of target detection and recognition. Additional redundant data are available, modelled here on prior point detection system performance.

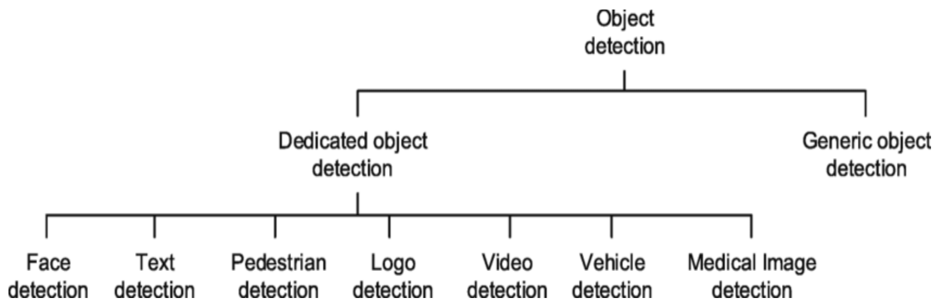

Fig. 1. Object detection has two kinds of applications: general object detection and specific object detection.

2. CONVOLUTIONAL NEURAL NETWORK (CNN)

CNN is a feed forward neural network for artificial intelligence. It is a popular image recognition tool. The input data is represented by CNN as a multidimensional array. It works well when there is a lot of labelled data. CNN extracts each part of the input image, which is called the receptive field. It adds weight to each neuron according to the importance of the receptive field. In this way, the relative importance of the distinguishing neuron can. The architecture of CNNis shown in the figure. Convolution layers, pool layers, and fully linked layers are the three types of layers that make up the CNN architecture. CNN has a wide range of applications in various fields. That convolution layer is the first layer. The basic building block of CNN is the convolution layer. It extracts high level functions in the input signal.

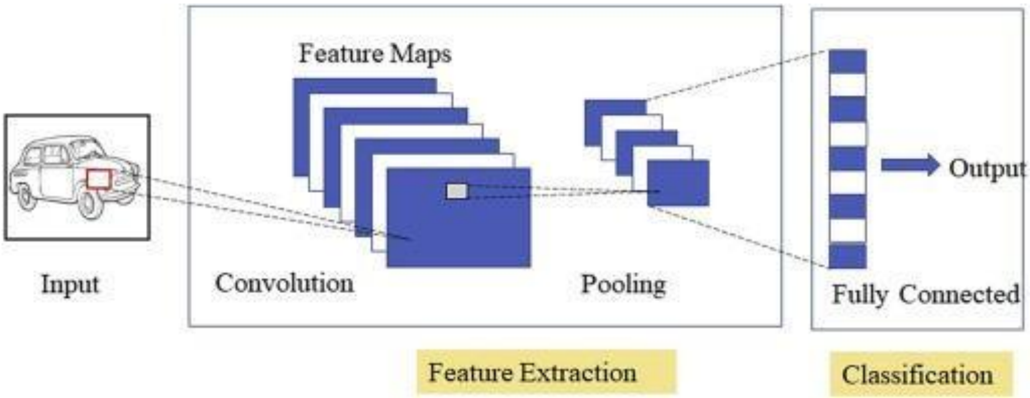

Fig. 2. CNN Layer architecture

After the convolution layer, a pool layer is applied. The pooling operation is set according to the application. Maximum pooling, minimum pooling and average pooling are three independent pooling operations. The merging operation is mainly used to reduce dimensionality and select the most important functions. The fully linked layer containing the activation function receives these attributes. The convolution operation using several filters can extract functions (function maps) those routines are used to maintain the data set and also the related spatial information. This operation collecting the function card first from convolution process is also known under sampling. The most often used grouping activities on CNN are maximum pooling & average pooling.

2.1. Convolutional Layer

The convolution layer consists of a set of convolution nuclei, where each neuron acts as a nucleus. Use the core of the convolution to divide the image into small parts, called thereceptive field. When the image is divided into small parts, it is easier to extract the characteristic pattern. By multiplying its components by the corresponding elements in the receptive field, the core uses a predefined set of weights to convolve with the image. Due to the weight sharing capability of the convolution operation, the same weight set can be applied to the image through the sliding core, so the CNN parameters are more efficient than the fullyconnected network. Depending on the type and size of the filter, the type of filling and the direction of convolution, the convolution operation can be further divided into different types.

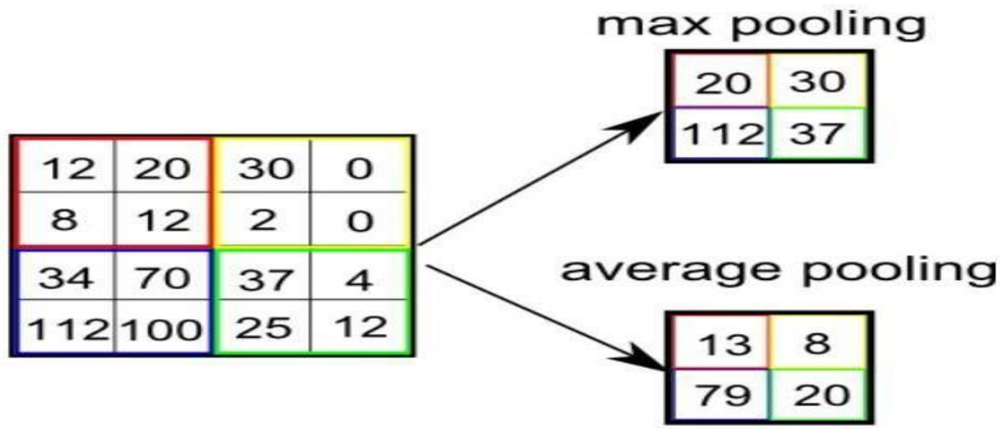

Fig.3 Pooling Operations

2.2. Pooling Layer

Just after convolution layer, the pool layer was generally applied. The primary aim of such a layer would be to reduce computer expense by reducing the size of the convolutionary function card. This one is done when the connections among layers are reduced and each function card operates independently. There are several pooling operations depending on the method employed. The biggest element is taken in maximum pooling from the function card. Just use average pooling in such an image segment of such a particular size to compute the average value of something like the components. In a predetermined segment, the sum pooling calculates overall total of components. The pooling layer provides a method for testing the Functions Card through listing features inside the function card patch.

2.3.Fully Connected Layer

The fully connected layer, a feed-in neural network, is relatively basic. The completely linked layer constitutes the network's last layers. The input of the entirely connected layer seems to be the output from the last layer, which itself is flattened out but then given to the completely connected layer. Neurons and weights create a completely connected layer of neurons (FC), used this to link the neurons between the two layers. Usually, this output layer precedes the final layers of the CNN architecture. The input pictures over the first few layers are combined in this stage and transferred to the FC layer. The flattened vector is then sent through several additional FC layers, which normally perform mathematical function operations. The classification process starts at this point.

3. TYPES OF OBJECT DETECTION ALGORITHMS

Algorithms available for object detection can be divided into two categories: classification- based algorithms and regression-based algorithms. 1) Classification based algorithms Classification based algorithms are implemented in two stages. The initial stage is the selection of region that is of interest (RoI) in the image. Then these regions are classified with the use of a convolutional neural network. This approach of performing one stage prior to the other can be slow due to the need to run the prediction algorithms on each region selected in the first stage. Few common examples for this type of algorithms are the Retina Net, Region-based CNN (RCNN), the Fast-RCNN, Faster R-CNN and Mask-RCNN (which is known to be a state-of-art under regional-based CNN algorithms). 2) Regression-based algorithms Regression-based algorithms are implemented so that instead of selecting and singling out regions of interest in an image, they predict classes and their relevant bounding boxes for the whole image in one run through the model. Since frame detection is treated as a regression problem, a complex pipeline is not necessary for regression-based algorithms. Famous examples of this type of algorithmsare the Single Shot Multibox Detector (SSD) and YOLO algorithms. Due to the simultaneousness of the detection and its nature of high speed (achieved with a tradeoff with accuracy), these are commonly used for real-time object detection. The detection and understanding of the more popular YOLO algorithms require an initial establishment of what will be predicted before the models are used. The prediction would result in a bounding box (specifying the Object's location) along with a class that has the highest probability amongst the established set of classes.

4. YOU ONLY LOOK ONCE (YOLO) ALGORITHM

YOLO is a novel approach to detect multiple objects present in an image in real-time while drawing bounding boxes around them. It passes the image through the CNN algorithm only once to get the output, thus the name. Although comparatively similar to R-CNN, YOLO practically runs a lot faster than Faster R-CNN because of its simpler architecture. Unlike Faster R-CNN, YOLO can classify and perform bounding box regression at the same time. With YOLO, the class label containing objects, their location can be predicted in one glance. Total difference between YOLO as well as the standard CNN pipeline is the problem of regression of object detection through a separation of spatial boundaries and associated class probabilities, predicted by such a single neural network. This process of performing both bounding box prediction and class probability calculations is a unified network architecture that YOLO initially introduced. YOLO algorithm extends GoogLeNet equations to be used as their base forwarding transport computation, assumedly the reason behind the speed and accuracy of YOLO's real-time object detection. In comparison with R-CNN architectures, unlike running a classifier on a potential bounding box, then re-evaluating probability scores, YOLO predicts bounding boxes and class probability for those bounding boxes simultaneously. This optimizes the YOLO algorithm and is one of the significant reasons why YOLO is so fast and less likelyto have errors to be utilizable for real-time object predictions. YOLO's architecture is similar toa typical convolutional neural network inspired by the GoogLeNet model for image classification. The first layer of the network extracts images features and predicts output possibilities & co-ordinates through the fully linked strata. The whole YOLO network modelhas been constructed using 24 convergence layers, two completely linked layers, 1x1 reduction layers and 3x3 convolutionary layers [12].

5. RCNN BASED ALGORITHMS

The development among editions in the R-CNN series of publications was generally in terms of computing efficiencies, minimization of test time, and improved performance (mAP). These systems typically include the following components: i) a region proposal technique to produce "bounding boxes" or areas of probable objects in the image; ii) a feature generation stage to obtain features, typically using a CNN; iii) a classifying layer to estimate the class of object; and iv) a regression layer to improve the precision of the bounding box coordinates. The architecture of R-CNN is the result of combining many methods. It starts by selecting 2000 region suggestions that may include items, using a selection search method. To produce feature maps, area suggestions, or Regions of Interest (RoIs), is passed through a convolutional net. A an SVM model for classification in which feature maps are passed and a regression model is used for obtaining tight bound-ing boxes. Although unique at the time, this approach is exceedingly sluggish. Fast R-CNN outperformed R-CNN. In this paradigm, the entire picture is sent to a single CNN rather than each of the 2000 regions being fed to individual CNNs. This produces a feature map that includes all of the regions of interest. A comparable approach to that employed in R-CNN is utilized to choose region suggestions. The feature maps of all region suggestions are extracted and resized to the same size using a RoI pooling layer. This is then sent to fully connected layers with two branches: a softmax classifier to provide probabilities for each class and a bounding box regressor to provide accurate bounding box coordinates. Until Faster R- CNN was released, its competitors used a variety of CPU-based methods for region proposal, causing a bottleneck. The selection search technique was replaced by a new network called the Region Proposal Network, which enhanced the object identification design (RPN). The remaining of the architecture is the same as Fast R-CNN: the picture is sent to a CNN to create afeature map, from which features for RPN-proposed areas are chosen and scaled by a pooling layer, and fed to an FC layer with two segments, a softmax classifier, and a bounding box regressor. The detecting speed was improved because of this approach.

6. LITERATURE SURVEY

(Median Hardiv Nugraha etal 2020) suggested that tourism has emerged as one of the most promising sources of income in some countries. Implementing information technology to attract a large number of tourists is one strategy for increasing revenue from the tourism sector. Smart tourism is a technology that can be utilized. A mobile VQA (Visual Issue Response) software that pro- vides extensive information on tourism assets that used a cell phone's cameras is one ofthe intelligent tourism implementations there in Indonesian region. This study aims to develop a model for training that has high accuracy in detection. The result of the model formation process relies on the items for such discovery model used for VQA. The dataset for this study consists of 600 Monas images and 25 classes of peripheral items. The methods used for object detection are YOLO and RetinaNet, which search and compare accuracy with metrics from the two approaches. Using the original dataset, YOLO has an average accuracy (mAP) score of 60.77 to 71.99%, reliability level thresholds range from 0.1 to 0.9, and RetinaNet has a mAP score of72.18 to 92.98. %is. Using the extended dataset, YOLO's mAP score range is 52.51% and 93.72%, and RetinaNet's mAP score is 23.8% and 56.19%. The original dataset has an AUC (Area Under Curve) score of 0.99, and the complemented dataset has an AUC score of 0.96 using the YOLO approach. "According to the evaluation results, YOLO can detect objects better than RetinaNet, and the complemented dataset can detect objects better than the original dataset."

(Reagan L. Galvez et.al 2019) There are difficult efforts for baggage inspectors to manually find dangerous goods on Xray equipment at airports, train stations and elsewhere. When X-ray equipment is inspecting baggage, rotating the baggage often hides objects, making it difficult to recognize. It is difficult to find, especially during rush hour. In this investigation, a You Only Look Once (YOLO) based object detector is introduced as a method to automatically detect a threatening object in an X-ray image. We compared transfer learning and training from scratch on an IEDXray dataset consisting of IED replica-scanned X-ray pictures. Detecting a threat object quickly takes precedence over YOLO training transfer learning in the first place. The average accuracy (mAP) of the 416416 images was 45.89%. For 608608 images, it was 51.48%. For multi-scale images, it was 52.40%. On the other hand, transfer learning achieved only 29.54% mAP, whereas multi-scale images achieved 29.54% mAP.

(Md. Bahar Ullah et.al 2020) This work describes CPU-based YOLO, a real-time object detection paradigm for non-GPU computers to support low-configuration PC users. When it comes to object retrieval, there are many improved algorithms such as YOLO, Faster RCNN, RCNN, Mask RCNN, RFCN, SSD, RetinaNet, etc. To detect objects, YOLO employs Deep Neural Network technology, which is much faster and more accurate than most other methods. To operate YOLO, you need a GPU-based PC with a graphics card of 12 GB or more. We have merged OpenCV and YOLO into our model to run real-time object identification on CPU-based systems. On many non-GPU work-stations, our model detects objects in the video at a rate of 10.12 16.29 frames/sec with 95% confidence. When I run YOLO from the CPU, the mAP reaches 31.

(Chun Ju Huang et.al 2020), resulting in the development of point cloud and image deployment methods. It is important to understand that you must first use depth information before performing object splitting. The RYOLO neural network is used for object detection to recognize objects based on the result of object splitting. "Our approach brings real-time LiDAR point clouds and im-age recognition results in real time. Build high-speed line chips that improve system performance using cell-based design flows".

(Yuanyuan Hu et.al 2019) The use of unmanned aerial vehicles (UAVs) makes modern life easier, but it has some drawbacks. The key to anti UAV is its object detection technology. Capability to capture deep and high-level characteristics YOLOv3 is one of the best single stage detectors. This is the first time in this study that we have improved YOLO v3, detected UAVs more accurately, and detected UAVs for anti UAVs using YOLO v3-based algorithms. Instead of predicting the bounding box of an object using only the last three scales of the paper's feature map, detecting the last four scales of the feature map provides more texture and contour information. To use. Based on the input data, estimate the UAV with four scale feature maps and change the number of anchor boxes accordingly to reduce calculation time. Experiments have shown that using the pro-posed approach, UAVs can be identified more accurately and UAV detection techniques can be used to obtain more accurate boundary boxes, which can be used in the anti UAV field. increase.

(Guohe Zhang et.al 2019) You only find once (Yolo), a real-time object detection system thatis in the cutting edge of the technology. Rocket, a RISCV core, which acts as a controller for Yolo hardware accelerator. For this accelerator, user-defined extended instructions were pro- posed on the basis of RISCV. After the Xilinx `Virtex7 FPGA VC709, after the completion of the Yolo algorithm, the accelerator takes about 400 milliseconds to complete, and it is expected to have a higher speed with additional calculation units.

(Chengji Liu et.al 2019) discovered deep learning objects that showed excellent results. There-fore, noise, blur and swirling jitter are popular in real world photos. Therefore, the detection of objects due to the result of the simulation of real world recording measures is worse, we have created image-based images based on Yolo networks and contain image processing methods standard photos. After establishing various recession models, we have examined the effects of different recession models on the object recognition. We used the Yolo network to train a powerful model to increase the accuracy of traffic signs.

(Reagan L. Galvez et al.) All published in his work [1], in this article, uses a convulsive neural network (CNN) to detect objects in the surrounding environment. Compare MobileNetV1's single-shot (SSD) multi-box detector and the Faster-RCNN regional convolutionary network with object detection with InceptionV2. The results reveal that one model is best suited for real- time applications because of its speed while the other model may be utilized to detect objects more precisely. There seems to be a balance between precision and speed based on the outcomes. Use SSD in con-junction with MobileNetV1 when they require rapid registration, especially in real-time applications. Faster-RCNN with InceptionV2 was recommended for highaccuracy detection. There seems to be a balance among accuracy and speed, as per research data. Use SSD in conjunction with MobileNetV1 when they require rapid registration, especially in real-time applications. Use faster RCNN with InceptionV2 for high accuracy detecting features.

(Dr. Ulagamuthalvi et al. 2019) have proposed a model [2] that records text from images. Use classifiers like SVM and neural networks to train the classifier to detect objects when presenting new im-ages. Use delimiters to detect and locate text. The model is trained using a neural network, and the training set consists of a large number of photos with text. The model's performance was evaluated and it was found that it detected text when a new image appeared. The alternative growth neural network for the classification is utilized as well as the model suggested is highly accurate.

(Ren Shaoqing et al.) [3] proposed the Regional Proposal Network (RPN) was launched to enable free regional proposals and to share the whole image convergence function with the detection net- work. Simultaneously. RPN has been taught to develop high quality regional suggestions for rapid R-CNN detection from start to end. By sharing their configuration characteristics, they merge RPN and Fast R-CNN into one network - the RPN component advises the united network where and how to lead using the popular neural network name and indeed the "watchfulness" procedure. This technique creates a unified, deeply based object detection system which operates almost in real-time. The learnt RPN also enhances the regional proposal's quality and therefore increases the total objective detection accuracy.

In the paper published by **(Guo Tianmei et al.)** [4], A basic neural network for categorization of the picture has been created. The classification of images is done to use a basic neural network of convolution. We used the benchmark datasets minist and cifar-10 in our test. We also studied alter-native techniques for learning rate sets and various optimization algorithms that use convolutional neural networks to obtain the best parameters that affect image classification. They also studied the effect of alternative methods of learning rate sets and various optimization algorithms for obtaining the best parameters using convolutional neural networks on image classification. They also con-firmed that superficial networks have significant recognition capabilities.

(Joseph Redmond et al.) [5] released YOLO, a new object detection technology that was demonstrated. In detection purposes, the classifier created for object detection gets reused. Rather, object recognition having geographically dispersed border boxes and the corresponding class probabilities is seen as a regression issue. In one assessment, the border box and class probability for the whole picture are directly predicted by a single neural network. Since the whole detection pipeline is one network, the detection performance through start to end can be improved immediately. The model was simple to build and could be directly trained over the whole photograph. YOLO gets taught in such a loss function which matches the detection performance directly. Unlike classification technology, the whole model is presented together.

For the detection of objects, (**Tanvir Ahmad et al.**) An improved neural network based on YOLOv1. al [6] in this paper. There in following areas, the latest product for the neural network was enhanced. The YOLOv1 network loss function is first changed. The updated model uses scale type instead of margin style. The new loss feature seems to be more versatile andacceptable than the previous one was in terms of optimizing network failures. Third, there is a beginning model with a 1x1 convolution core that decreases the weight parameter again for number of layers. The construction model structure had also been incorporated. The results of both the network proposed were compared to the results of R-CNN and YOLOv1 whichdemonstrated the efficiency of the approach provided.

(Xin Zhang) recently published a paper [7] discussing the development, improvement, and weakness of the two-stage measurement detection algorithm represented by the RCNN series and single-phase target detection technology represented by the Yolo series. The object detection algorithm is faster than the two-stage object detection algorithm, and the two-stage object detection algorithm is more accurate than the first-level object detection algorithm. Target detection technology based on deep learning is better than classic detection methods in terms of accuracy and real-time performance. However, due to the complexity and diversity of real-world scenarios, many problems still exist. How to reduce the influence of complex background on target detection, and how to reduce the loss of accuracy caused by the change in target size and shape, has become a hot topic in this area.

This paper [8] was published by (**Pranav Adarsh et al)**. The article provides an overview of the methods for object recognition through covering 2 kinds of object detectors. Among two step detector algorithms are “RCNN, Fast RCNN and Faster RCNN, whereas YOLO v1, V2, V3 and SSD” are covered within the one-stage detector. Accuracy seems to be more important in the second level detector whereas speed is indeed the major focus of the very first level detector. An enhanced version of YOLO dubbed YOLO v3-Tiny will just be explained as well as its comparison with prior object detection and identification methods will indeed be described. YOLO v3- Tiny uses a pool layer to reduce the number of convolution layers. It generates a three-dimensional tensor with two scales of object score, boundary box and category prediction. An image is divided into SS lattice cells. We will ignore that the objectivity score is not the best boundary box for the final disclosure. YOLO v3-Tiny's feed forward architecture uses convolution layers and maximum pool layers to extract features.

(Zhitian et al.) [9] Proposed a model for the technique to be solved per. Pixel object detection prediction comparable to semantic segmentation and just a complete one-step object detector created (FCOS). Almost all the most advanced object detectors, such as Retina Net, SSD, YOLOv3 and Faster R-CNN, are supported by predefined anchor boxes. The FCOS detector we recommend is free and for free. When removing the pre-set anchor box set, FCOS avoids the complicated calculation of the anchor box e.g., Overlap calculation during training. In addition, all the hyperparameters of the anchor box, which are very sensitive to the final detection performance, are avoided. FCOS can also provide the latest one-step detection service. FCOS can be used as an RPN in the two-stage faster R-CNN detector, which is much higher than its RPN. Joseph Redmond et al. al [10] introduced some improvements of YOLO in this paper. YOLO struggles to spot small objects. But with YOLOv3 we can see performance improvementof small objects because the shortcut connection YOLOv3 uses a new network to extract functions. The new system is a hybrid method for YOLOv2 (Darknet-19) and the rest of the network. It has 53 overwhelming layers, called Darknet-53. Compared to YOLO and YOLO2, which predict output from the last scale, YOLOv3 predicts boxes with 3 different scales. YOLOv3 does not use soft-max, but uses a separate logistics classifier for each category.

(Ross Girshick) et al. al [11] proposed a method that integrates two key insights: (1) the use of

CNN to locate and segment objects in bottom-up regions can be achieved with high capacity, and (2) by labelling training When data is limited, monitored prior to training is used for auxiliary tasks, followed by specific field refinement, yielding significant results. This technique is called R- CNN: CNN Feature Region. This is the first research that shows that CNN can take the lead. Compared to simple systems based on HOG-like features, PASCAL VOC detection performance can be significantly improved. To do this, we focus on two issues: identifying objects with deep networks and training high-capacity models, with some notes on detection.

(Ross Girshick) [12] published another article in which he proposed a region-based method for rap-id target recognition (Fast R-CNN). Fast R-CNN relies on past work to effectively classify proposals with deep convolutional networks. Fast R-CNN combines various innovations with its previous work to increase the speed of training and testing, while improving detection accuracy. Spatial Pyramid Pooling Networks (SPPnets) make it possible to accelerate R-CNN through shared computing.

7. CONCLUSION

This research examines several object detections, tracking and recognition methods as well as function descriptions and segmentation methods based on video frames and tracking techniques. With new concepts, this method is used to improve object detection. The first-level object detection algorithm is faster than the second-level object detection algorithm, As well as the object- detection second-level method seems to be more precise than the object-detection first-level approach. Deep-learning target detection technology offers better in terms of accuracy and then in terms of performance than conventional detection approaches. A novel single-stage model technique has been found which could also enhance speed with loss of precision. The findings of the comparison show that now the YOLO method increases the object sensing speed while maintaining precision. The model is simple to construct and can be trained directly on the entire image, while the two-step model has a complex structure and is difficult to train.

REFERENCES

[1] IEEE Region 10, IEEE Seoul Section, Institute of Electrical and Electronics Engineers. Korea Council, and Institute of Electrical and Electronics Engineers, “Proceedings of TENCON 2018 : 2018 IEEE Region 10 Conference : Object Detection using Convolu-tional Neural Network : Jeju, Korea, 28-31 October 2018,” *TENCON 2018 - 2018 IEEEReg. 10 Conf.*, no. October, pp. 2023–2027, 2018.

[2] Ulagamuthalvi, J. B. J. Felicita, and D. Abinaya, “An efficient object detection modelusing convolution neural networks,” *Proc. Int. Conf. Trends Electron. Informatics, ICOEI 2019*, no. Icoei, pp. 142–147, 2019, doi: 10.1109/ICOEI.2019.8862698.

[3] S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, 2017, doi: 10.1109/TPAMI.2016.2577031.

[4] Institute of Electrical and Electronics Engineers, “2017 IEEE 2nd International Con- ference on Big Data Analysis (ICBDA 2017) : March 10-12, 2017, Beijing, China.,” pp. 721–724, 2017.

[5] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only look once: Unified, real-time object detection,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2016-Decem, pp. 779–788, 2016, doi: 10.1109/CVPR.2016.91.

[6] T. Ahmad *et al.*, “Object Detection through Modified YOLO Neural Network,” *Sci. Program.*, vol. 2020, no. June, 2020, doi: 10.1155/2020/8403262.

[7] J. Lv, B. Zhang, and X. Li, “Research on object detection algorithm based on PVANet,”*Adv. Intell.Syst. Comput.*, vol. 759, no. 4, pp. 141–151, 2019, doi: 10.1007/978-981- 13-0341-8_13.

[8] P. Adarsh, P. Rathi, and M. Kumar, “YOLO v3-Tiny: Object Detection and Recogni- tion using one stage improved model,” *2020 6th Int. Conf. Adv. Comput. Commun. Syst. ICACCS 2020*, pp. 687–694, 2020, doi: 10.1109/ICACCS48705.2020.9074315.

[9] Z. Tian, C. Shen, H. Chen, and T. He, “FCOS: Fully convolutional one-stage object detection,” *Proc.*

IEEE Int. Conf. Comput. Vis., vol. 2019-Octob, no. Iccv, pp. 9626– 9635, 2019, doi: 10.1109/ICCV.2019.00972.
[10] J. Redmon and A. Farhadi, "YOLO v.3," *Tech Rep.*, pp. 1–6, 2018, [Online]. Available: https://pjreddie.com/media/files/papers/YOLOv3.pdf.
[11] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurateobject detection and semantic segmentation," *Proc. IEEE Comput. Soc. Conf. Comput.Vis. Pattern Recognit.*, pp. 580–587, 2014, doi: 10.1109/CVPR.2014.81.
[12] R. Girshick, "Fast R-CNN," *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2015 Inter, pp. 1440–1448, 2015, doi: 10.1109/ICCV.2015.169.
[13] Median Hardiv Nugraha;Dina Chahyati Tourism Object Detection Around MonumenNasional (Monas) Using YOLO and RetinaNet 2020 International Conference on Ad-vanced ComputerScience and Information Systems (ICACSIS) Year: 2020
[14] Reagan L. Galvez;Elmer P. Dadios;Argel A. Bandala;Ryan Rhay P. Vicerra YOLO- based Threat Object Detection in X-ray Images 2019 IEEE 11th International Confer-ence on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM) Year: 2019
[15] Md. Bahar Ullah CPU Based YOLO: A Real Time Object Detection Algorithm 2020 IEEE Region10 Symposium (TENSYMP) Year: 2020
[16] Chun Ju Huang;Ting-Wei Chen;Yu-Cheng Fan Chip and System Design of Real TimeObjectDetection Based on LS-R-YOLO Network 2020 IEEE 9th Global Conference on Consumer Electronics (GCCE) Year: 2020
[17] Yuanyuan Hu;Xinjian Wu;Guangdi Zheng;Xiaofei Liu Object Detection of UAV for Anti-UAVBased on Improved YOLO v3 2019 Chinese Control Conference (CCC) Year: 2019
[18] Guohe Zhang;Kepeng Zhao;Bin Wu;Yiqun Sun;Li Sun;Feng Liang A RISC-V based hardware accelerator designed for Yolo object detection system 2019 IEEE Interna- tional Conference of Intelligent Applied Systems on Engineering (ICIASE) Year: 2019
[19] Chengji Liu;Yufan Tao;Jiawei Liang;Kai Li;Yihang Chen Object Detection Based onYOLO Network 2018 IEEE 4th Information Technology and Mechatronics Engineer-ing Conference (ITOEC) Year: 2018

AUTHORS

Utkarsh Namdev
Utkarsh Namdev is a final year student of Dual degree Integrated P.G. programme (BE+M.TECH) in Department of Computer Science & Engineering at University Institute of Technology, Rajiv Gandhi Proudyogiki Vishwavidyalaya, Bhopal (MP) India.

Shikha Agrawal
Dr. Shikha Agrawal is an Associate Professor in Department of Computer Science & Engineering at University Institute of Technology, Rajiv Gandhi Proudyogiki Vishwavidyalaya, Bhopal (MP) India.

Rajeev Pandey
Dr. Rajeev Pandey is an Associate Professor in Department of Computer Science & Engineering at University Institute of Technology, Rajiv Gandhi Proudyogiki Vishwavidyalaya, Bhopal (MP) India.