Component Detection with Deep Learning A review

Harshal Jorigal[1], Shivani Thorat[1], Karuna Rathod[1], Nikita Doiphode[1], Mrs. Geeta Atkar[2]

[1] *UG Students,* [2] *Professors, Computer Engineering Department, G H Raisoni College of Engineering and Management, Pune*

ABSTRACT

Object detection is a computerized technology-related Computer vision and image processing. It is a technique that allows us to identify and locate objects in an image or video. With the help of this technology the identification and localization of objects or any component can be simplified, object detection can be used to count objects in a scene and determine and track their precise locations, all while accurately labeling them. Object detection algorithms typically leverage machine learning or deep learning to produce meaningful results. The goal of object detection is to replicate this intelligence using a computer. Object detection can be done more accurately using certain algorithms. OpenCV which is a python library that functions mainly on computer vision, Yolo is the most efficient real-time algorithm which is used to effectively detect an object. In deep learning, CNN TensorFlow Keras is the algorithms used to accurately locate and identify objects.

Keywords— computer vision, OpenCv, Tensorflow, Keras, yolo, CNN.

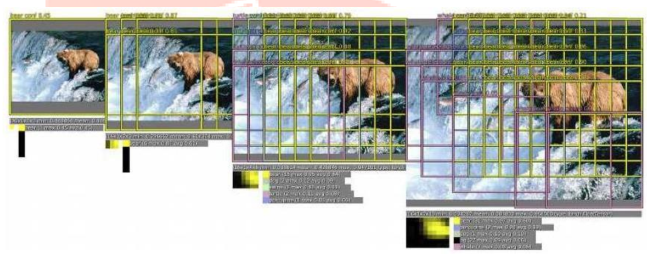
I. INTRODUCTION

Object detection is a technique that allows to identify and locate objects present either in the real world or in an image or video where the object can be anything such as humans, machines, etc.it provides various features to identify Objects. humans can easily recognize any object without any efforts in the real world by themselves but for machines, it cannot be possible to recognize objects or things. Object detection is totally linked to other computer vision techniques such as image recognition and image segmentation. It helps us to understand and analyze things in images or videos. It is mostly used in image annotations, activity recognition, face detection, face recognition, image retrieval, security surveillance, machine inspection, automated vehicle systems, etc. It is also used in tracking things, for example tracking movements of a person, etc. Basically, it deals with finding and locating specific objects. the process of locating moving objects using the camera in video sequences over time is known as object tracking. the main aim of object tracking is to relate target objects in consecutive video frames. Object tracking requires features of objects or the shape of the object and location. To detect or locate the moving object in the frame

II. LITERATURE SURVEY

This section of the literature survey eventually reveals some facts based on thoughtful analysis of many authors work as follows.

-OverFeat: Integrated Recognition, Localization and Detection using Convolutional Networks[1]

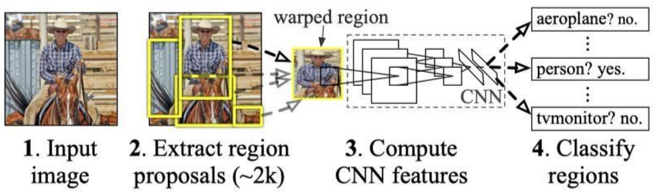


Localization/Detection pipeline
Figure no.(1)[1]

Inspired via way of means of the early fulfillment of AlexNet withinside the 2012 ImageNet competition, wherein CNN-primarily based totally characteristic extraction defeated all home made characteristic extractors, OverFeat fast delivered CNN lower back into the item detection place as well. The concept may be very straightforward: if we will classify one picture the usage of CNN, what approximately greedily scrolling thru the complete picture with one of a kind sizes of windows, and try and regress and classify them one-via way of means of-one the usage of a CNN? This leverages the energy of CNN for characteristic extraction and category, and additionally bypassed the tough area notion trouble via way of means of pre-described sliding windows. Also, for the reason that a close-by convolution kernel can percentage a part of the computation result, it isn't always important to compute convolutions for the overlapping place, as a result lowering value a lot. OverFeat is a pioneer withinside the one-degree item detector. It attempted to mix characteristic extraction, vicinity regression, and area category withinside the equal

CNN. Unfortunately, this sort of one-degree method additionally suffers from fantastically poorer accuracy because of much less previous information used. Thus, OverFeat did not lead a hype for one-degree detector research, till a far extra stylish answer popping out 2 years later.

-Region-based Convolutional Networks for Accurate Object Detection and Segmentation[2]

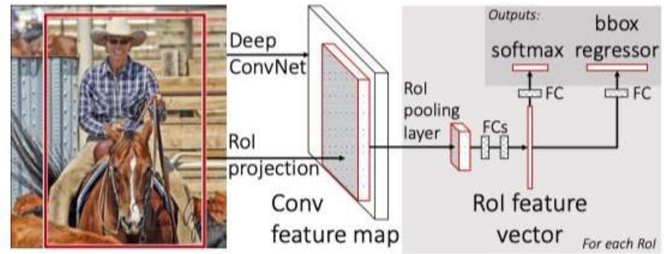


R-CNN: Region-based Convolutional Network
Figure no.(2)[2]

Also proposed in 2013, R-CNN is a chunk past due as compared with OverFeat. However, this vicinity-primarily based totally method sooner or later caused a huge wave of item detection studies with its two-level framework, i.e, vicinity notion level, and vicinity type and refinement level.

In the above diagram, R-CNN first extracts capacity areas of the hobby from an enter photograph with the aid of using the usage of a way referred to as selective seek. Selective seek doesn't absolutely try and recognize the foreground item, instead, it businesses comparable pixels with the aid of using counting on a heuristic: comparable pixels normally belong to the equal item. Therefore, the consequences of selective seek have a completely excessive opportunity to incorporate something meaningful. Next, R-CNN warps those place proposals into fixed-length pics with a few paddings and feeds those pics into the second level of the community for extra fine-grained recognition. Unlike the one's vintage techniques the usage of selective seek, R-CNN changed HOG with a CNN to extract functions from all place proposals in its 2d level. One caveat of this technique is that many place proposals aren't absolutely a complete item, so R-CNN desires to now no longer handiest discover ways to classify the proper classes, however additionally discover ways to reject the poor ones. To remedy this problem, R-CNN dealt with all place proposals with a ≥ 0.five IoU overlap with a ground-reality field as positive, and the relaxation as negatives.
Region notion from selective seek incredibly relies upon the similarity assumption, so it could handiest offer a difficult estimate of location. To in addition enhance localization accuracy, R-CNN borrowed a concept from "Deep Neural Networks for Object Detection" (aka DetectorNet) and added an extra bounding field regression to expect the middle coordinates, width, and peak of a field. This regressor is broadly used withinside the destiny item detectors.
However, a -level detector like R-CNN suffers from massive issues: 1) It's now no longer absolutely convolutional due to the fact selective seek isn't E2E trainable. 2) place notion level is normally very sluggish in comparison with different one-level detectors like OverFeat, and going for walks on every place notion one by one makes it even slower. Later, we can see how R-CNN evolve through the years to cope with those issues.

-Fast R-CNN[3]

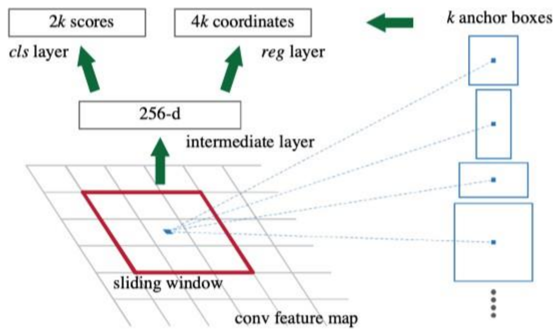


Fast R-CNN architecture
Figure no.(3)[3]

A brief follow-up for R-CNN is to lessen the reproduction convolution over more than one vicinity proposal. Since those vicinity proposals all come from one picture, it's clear to enhance R-CNN with the aid of using going for walks CNN over the complete picture as soon as and proportion the computation amongst many vicinity proposals. However, extraordinary vicinity proposals have extraordinary sizes, which additionally bring about extraordinary output characteristic map sizes if we're the use the identical CNN characteristic extractor. These characteristic maps with numerous sizes will save you from the use of completely related layers for similar class and regression due to the fact the FC layer handiest works with a hard and fast length enter.
Fortunately, a paper called "Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition" has already solved the dynamic scale difficulty for FC layers. In SPPNet, a characteristic pyramid pooling is added among convolution layers and FC layers to create a bag-of-phrases fashion of the characteristic vector. This vector has a hard and fast length and encodes functions from extraordinary scales, so our convolution layers can now take any length of pix as entering without stressful approximately the incompatibility of the FC layer. Inspired with the aid of using this, Fast R-CNN proposed a comparable layer name the ROI Pooling layer. This pooling layer downsamples characteristic maps with extraordinary sizes into a hard and fast-length vector. By doing so, we are able to now use the identical FC layers for class and field regression, regardless of how huge or small the ROI is.
With a shared characteristic extractor and the scale-invariant ROI pooling layer, Fast R-CNN can attain a comparable localization accuracy however having 10~20x quicker education and 100~200x quicker inference. The close to real-time inference and a less difficult E2E education protocol for the detection component make Fast R-CNN a famous preference withinside the enterprise as well.

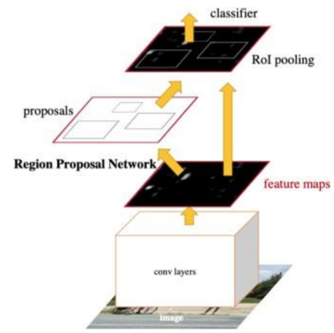
-Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks[6]

As we brought above, in early 2015, Ross Girshick proposed an progressed model of R-CNN referred to as Fast R-CNN with the aid of using the usage of a shared characteristic extractor for proposed regions. Just some months later, Ross and his crew got here again with any other development again. This new community Faster R-CNN isn't best quicker than preceding variations however additionally marks a milestone for item detection with a deep gaining knowledge of method.



Region Proposal Network (RPN)
Figure no.(4)[6]

With Fast R-CNN, the simplest non-convolutional piece of the community is the selectively seek location thought. As of 2015, researchers began out to recognize that the deep neural community is so magical, that it could analyze something given sufficient data. So, is it viable to additionally teach a neural community to thought regions, in preference to counting on heuristic and homemade methods like selective seek? Faster R-CNN observed this path and thinking, and efficiently created the Region Proposal Network (RPN). To honestly put, RPN is a CNN that takes a photo as entering and outputs a hard and fast of square item proposals, every with an objectiveness score. The paper used VGG at the beginning however different spine networks which include ResNet grow to be extra big later. To generate location proposals, a 3x3 sliding window is carried out over the CNN function map output to generate 2 scores (foreground and background) and four coordinates for every location. In practice, this sliding window is carried out with a 3x3 convolution kernel with a 1x1 convolution kernel.
Although the sliding window has a hard and fast size, our items may also seem on unique scales. Therefore, Faster R-CNN brought a way referred to as anchor field. Anchor containers are pre-described earlier containers with unique element ratios and sizes however proportion the equal significant location. In Faster R-CNN there are k=nine anchors for every sliding window location, which covers three-element ratios for three scales every. These repeated anchor containers over unique scales carry first-class translation-invariance and scale-invariance functions to the community whilst sharing outputs of the equal function map. Note that the bounding field regression may be computed from those anchor fields in preference to the entire photo.

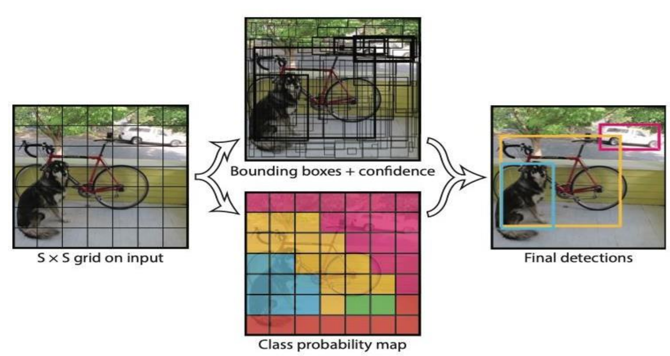


R-CNN As A Single Module
Figure no.(5)[6]

So far, we mentioned the brand new Region Proposal Network to update the vintage selective seek area proposal. To make the very last detection, Faster R-CNN makes use of the equal detection head from Fast R-CNN to do type and fine-grained localization. Do you recollect that Fast R-CNN additionally makes use of a shared CNN characteristic extractor? Now that RPN itself is likewise a characteristic extraction CNN, we will simply proportion it with a detection head just like the diagram above. This sharing layout doesn't deliver a few hassles though. If we educate RPN and Fast R-CNN detector together, we are able to deal with RPN proposals as a steady enter of ROI pooling, and unavoidably forget about the gradients of RPN's bounding container proposals. One stroll round is referred to as opportunity education in which you educate RPN and Fast R-CNN in turns. And later in a paper "Instance-conscious semantic segmentation through multi-undertaking community cascades", we will see that the ROI pooling layer also can be made differentiable w.r.t. the container coordinates proposals.

-*You Only Look Once: Unified, Real-Time Object Detection[4]*

While the R-CNN collection began out a huge hype over-degree item detection withinside the studies community, its complex implementation delivered many complications for engineers who preserve it. Does item detection want to be so cumbersome? If we're inclined to sacrifice a chunk of accuracy, are we able to alternate for a whole lot quicker speed? With those questions, Joseph Redmon submitted a community known as YOLO to arxiv.org handiest 4 days after Faster R-CNN's submission and subsequently delivered recognition again to one-degree item detection years after OverFeat's debut.

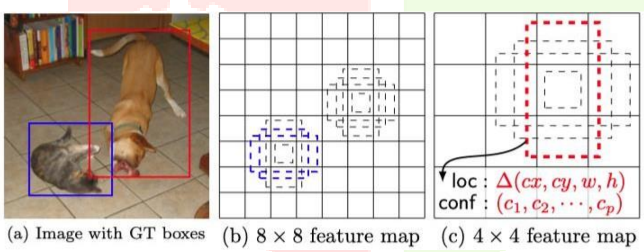


The Model
Figure no.(6)[4]

Unlike R-CNN, YOLO determined to address location concept and location class collectively withinside the equal CNN. In different words, it treats item detection as a regression hassle, as opposed to a class hassle counting on location proposals. The trendy concept is to cut up the entry into an SxS grid and having every mobileular immediately regress the bounding field area and the self-assurance rating if the item middle falls into that mobileular. Because items may also have distinctive sizes, there might be a couple of bounding field regressor in keeping with mobileular. During training, the regressor with the best IOU might be assigned to evaluate with the ground-fact label, so regressors on the equal area will discover ways to deal with distinctive scales over time. In the meantime, every mobileular may also are expecting C magnificence probabilities, conditioned at the grid mobileular containing an item (excessive self-assurance rating). This method is later defined as dense predictions due to the fact YOLO attempted to are expecting lessons and bounding containers for all viable places in an image. In contrast, R-CNN is based on location proposals to filter history regions, therefore the very last predictions are tons extra sparse.

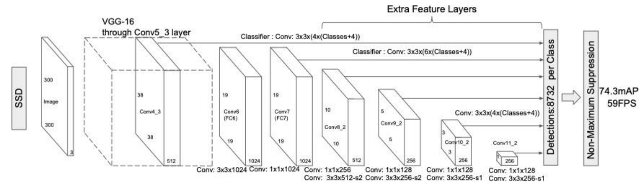
-*SSD: Single Shot MultiBox Detector[5]*

YOLO v1 established the potentials of one-degree detection, however, the overall performance hole from two-degree detection remains noticeable. In YOLO v1, a couple of items may be assigned to the equal grid cell. This became a huge mission whilst detecting small items, and has become an essential hassle to remedy as a way to enhance a one-degree detector's overall performance to be on par with two-degree detectors. SSD is any such challenger and assaults this hassle from 3 angles.



SSD Framework
Figure no.(7)[5]

First, the anchor container approach from Faster R-CNN can alleviate this problem. Objects withinside the equal place generally include one-of-a-kind factor ratios to be visible. Introducing anchor containers now no longer most effective improved the quantity of the item to locate for every cell, however additionally helped the community to higher differentiate overlapping small gadgets with this factor ratio assumption.



Single Shot Detection Model
Figure no.(8)[5]

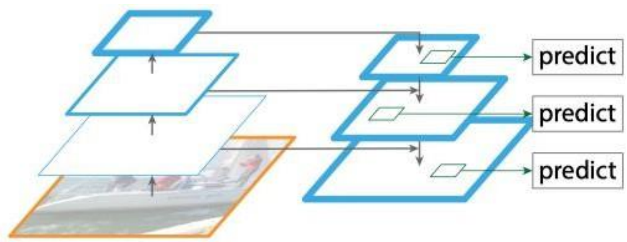
SSD went down in this avenue similarly via way of means of aggregating multi-scale capabilities earlier than detection. This is a totally not unusual place technique to choose up fine-grained neighborhood capabilities whilst keeping coarse international capabilities in CNN. For example, FCN, the pioneer of CNN semantic segmentation, additionally merged capabilities from a couple of degrees to refine the segmentation boundary. Besides, multi-scale characteristic aggregation may be without problems accomplished on all, not unusual place class networks, so it's very handy to switch out the spine with every other network.

Finally, SSD leveraged a big quantity of records augmentation, particularly centered on small objects. For example, snapshots are randomly increased to a mile's large length earlier than random cropping, which brings a zoom-out impact to the schooling records to simulate small objects. Also, big bounding packing containers are typically smooth to learn. To keep away from those smooth examples dominating the loss, SSD followed a tough bad mining method to choose examples with the very best loss for every anchor box.SSD went down in this avenue similarly via way of means of aggregating multi-scale capabilities earlier than detection. This is a totally not unusualplace technique to choose up fine-grained neighborhood capabilities whilst keeping coarse international capabilities in CNN. For example, FCN, the pioneer of CNN semantic segmentation, additionally merged capabilities from a couple of degrees to refine the segmentation boundary. Besides, multi-scale characteristic aggregation may be without problems accomplished on all not unusualplace class networks, so it's very handy to switch out the spine with every other network.

Finally, SSD leveraged a big quantity of records augmentation, particularly centered to small objects. For example, snap shots are randomly increased to a miles large length earlier than random cropping, which brings a zoom-out impact to the schooling records to simulate small objects. Also, big bounding packing containers are typically smooth to learn. To keep away from those smooth examples dominating the loss, SSD followed a tough bad mining method to choose examples with the very best loss for every anchor box.

-*Feature Pyramid Networks for Object Detection[7][19]*

With the release of Faster-RCNN, YOLO, and SSD in 2015, it looks like the overall shape of an item detector is determined. Researchers begin to examine enhancing every person a part of those networks. Feature Pyramid Networks is an try and enhance the detection head through the usage of functions from extraordinary layers to shape a function pyramid. This function pyramid concept isn't very novel in pc imaginative and prescient research. Back then while functions are nevertheless manually designed, a function pyramid is already a completely powerful manner to apprehend styles at extraordinary scales. Using the Feature Pyramid in deep mastering is likewise now no longer a brand new concept: SSPNet, FCN, and SSD all tested the advantage of aggregating multiple-layer functions earlier than classification. However, the way to percentage the function pyramid among RPN and the region-primarily based totally detector remains but to be determined.



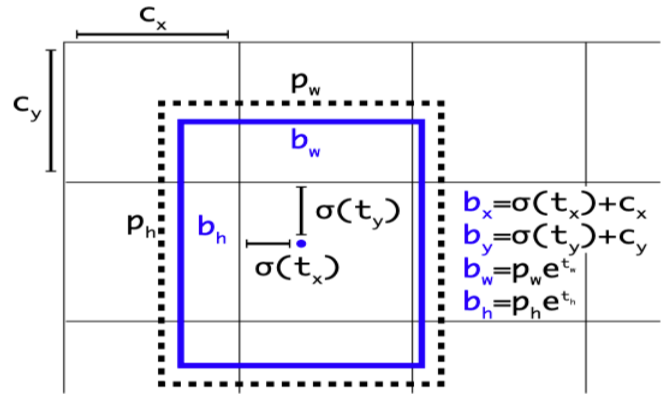
Pyramidal Feature Hierarchy
Figure no.(9)[7]

First, to rebuild RPN with an FPN shape just like the diagram above, we want to have an area inspiration walking on more than one extraordinary scale of characteristic output. Also, we most effectively want three anchors with extraordinary issue ratios according to the region now due to the fact items with extraordinary sizes could be taken care of with the aid of using extraordinary tiers of the characteristic pyramid. Next, to apply an FPN shape withinside the Fast R-CNN detector, we additionally want to conform it to locate on more than one scale of characteristic maps as well. Since area proposals may have extraordinary scales too, we ought to use them withinside the corresponding stage of FPN as well. In short, if Faster R-CNN is a couple of RPN and area-primarily based totally detector walking on one scale, FPN converts it into more than one parallel branch walking on extraordinary scales and collects the very last effects from all branches withinside the end.

-*YOLO9000: Better, Faster, Stronger[9]*

While Kaiming He, Ross Girshick, and their group hold enhancing their -level R-CNN detectors, Joseph Redmon, on the alternative hand, became additionally busy enhancing his one-level YOLO detector. The preliminary model of YOLO suffers from many shortcomings: predictions primarily based totally on a rough grid added decrease localization accuracy, scale-agnostic regressors in line with grid mobileular additionally made it hard to apprehend small packed objects. Fortunately, we noticed too many wonderful improvements in 2015 in lots of laptop imaginative and prescient areas. YOLO v2 simply desires to discover a manner to combine all of them to grow to be better, faster, and stronger. Here are a few highlights of the modifications:

- YOLO v2 introduced Batch Normalization layers from a paper called "Batch Normalization: Accelerating Deep Network Training through Reducing Internal Covariate Shift".



Bounding Boxes With Dimension Priors and Location Prediction
Figure no.(10)[9]

- Just like SSD, YOLO v2 additionally added Faster R-CNN's concept of anchor bins for bounding container regression. But YOLO v2 did a few customizations for its anchor bins. Instead of predicting offsets to anchor bins, YOLOv2 constraints the item middle regression tx and ty inside the accountable grid mobileular to stabilize early education. Also, anchors sizes are decided with the aid of using a K-way clustering of the goal dataset to higher align with item shapes.

- A new spine community known as Darknet is used for characteristic extraction. This is stimulated with the aid of using "Network in Network" and GooLeNet's bottleneck structure.

- To enhance the detection of small objects, YOLO v2 delivered a passthrough layer to merge functions from an early layer. This element may be visible as a simplified model of SSD.

- Last however now no longer least, Joseph found out that enter decision is a silver bullet for small item detection. It now no longer simplest doubled the enter for the spine to 448x448 from 224x224 however additionally invented a multi-scale education schema, which entails one-of-a-kind enter resolutions at one-of-a-kind durations of education.

Note that YOLO v2 additionally experimented with a model that's skilled on 9000 magnificence hierarchical datasets, which additionally represents an early trial of multi-label type in an item detector.

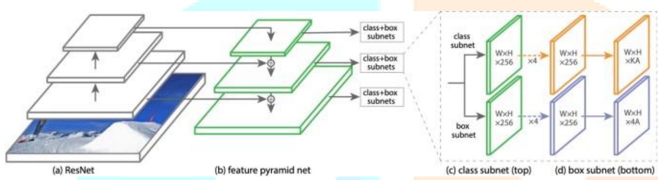
-*Focal Loss for Dense Object Detection[8]*

To apprehend why one-degree detectors are generally now no longer as exact as two-degree detectors, RetinaNet investigated the foreground-heritage elegance imbalance difficulty from a one-degree detector's dense predictions. Take YOLO as an instance, it attempted to expect instructions and bounding containers for all feasible places withinside the meantime, so a maximum of the outputs is matched to terrible elegance all through training. SSD addressed this difficulty via way of means online tough instance mining. YOLO used an objectiveness rating to implicitly educate a foreground classifier withinside the early degree of training. RetinaNet thinks they each didn't get the important thing to the problem, so it invented a brand new loss feature referred to as Focal Loss to assist the community study what's important.

$$\mathrm{FL}(p_t) = -\alpha_t(1-p_t)^{\gamma}\log(p_t).$$

Focal Loss Definition
Figure no.(11)[8]

Focal Loss brought a strength γ (they name it focusing parameter) to Cross-Entropy loss. Naturally, because the self-belief rating turns higher, the loss fee becomes a lot decrease than an everyday Cross-Entropy. The α parameter is used to stability this sort of focusing effect.

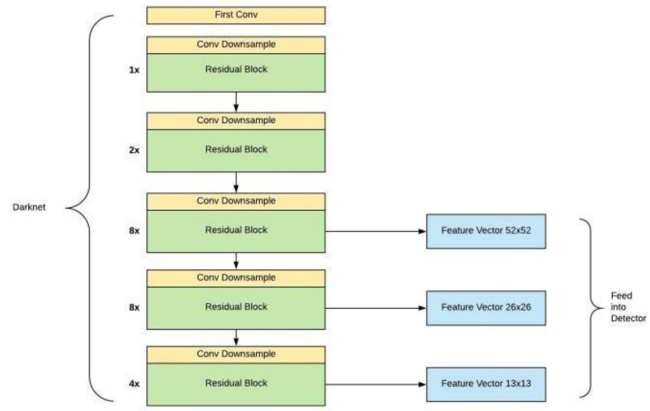


Ratina Net
Figure no.(12)[8]

This concept is so easy that even a number one college pupil can understand. So too in addition to justify their work, they tailored the FPN version they formerly proposed and created a brand new one-level detector known as RetinaNet. It consists of a ResNet spine, an FPN detection neck to channel capabilities at extraordinary scales, and subnets for type and field regression as detection head. Similar to SSD and YOLO v3, RetinaNet makes use of anchor bins to cowl objectives of numerous scales and component ratios.
A little bit of a digression, RetinaNet used the COCO accuracy from a ResNeXT-one zero one and 800 enter decision version to comparison YOLO v2, which handiest has a light-weighted Darknet-19 spine and 448 enter decision. This insincerity indicates the team's emphasis on getting higher benchmark results, as opposed to fixing sensible trouble like a speed-accuracy trade-off. And it is probably a part of the cause that RetinaNet didn't take off after its release.

-*Dive Really Deep into YOLO v3: A Beginner's Guide[11]*

YOLO v3 is the closing model of the reputable YOLO series. Following YOLO v2's tradition, YOLO v3 borrowed greater thoughts from preceding studies and was given a very effective one-level detector like a monster. YOLO v3 balanced the velocity, accuracy, and implementation complexity quite well. And it was given honestly famous withinside the enterprise due to its speedy velocity and easy components. If you're interested, I wrote a completely distinctive rationalization of ways YOLO v3 works in my preceding article "Dive Really Deep into YOLO v3: A Beginner's Guide".



YOLO Stack
Figure no.(13)[11]

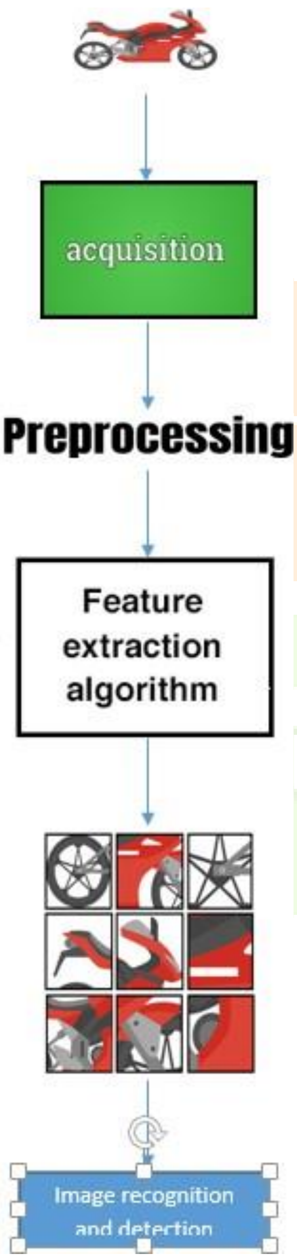
Flow
Figure no.(14)[11]

Simply put, YOLO v3's achievement comes from its extra effective spine function extractor and a RetinaNet-like detection head with an FPN neck. The new spine community Darknet-fifty three leveraged ResNet's pass connections to gain an accuracy that's on par with ResNet-50 however a whole lot faster. Also, YOLO v3 ditched v2's skip thru layers and completely embraced FPN's multi-scale predictions design. Since then, YOLO v3 eventually reversed people's influence of its bad overall performance whilst handling small objects.

Besides, there are some laugh records approximately YOLO v3. It dissed the COCO mAP 0.5:0. ninety-five metric, and additionally proven the uselessness of Focal Loss whilst the use of a conditioned dense prediction. The creator Joseph even determined to end the entire laptop imaginative and prescient studies 12 months later, due to his subject of navy usage.

III. PROPOSED ARCHITECTURE

Render Sample
Figure no.(15)

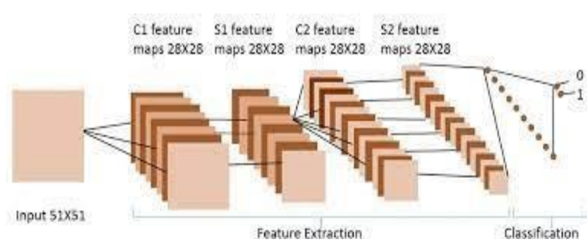


Flow Of System
Figure no.(16)

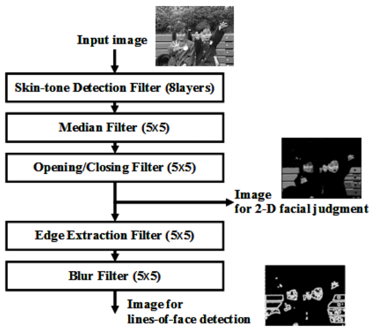
In the above figure the Object detection algorithms use features which can be extracted to recognize a particular object. This model is not at all complex, It is easy to implement. Here, object detection is a single regression problem which detects directly from bounding box coordinates and class probability. Every object has its own class such as all circles are round, which are used while recognizing the objects.

- Preprocessing-
  Preprocessing is the lowest level of abstraction. This process improves the image intensity by suppressing the features which we don't want for further processing .It resizes the image size to 448*448 and also normalizes the contrast and brightness effects. The image is also cropped and resized by which we can perform feature extraction easily. The input images are pre-processed and very easily normalize the contrasts and brightness. Preprocessing step can be done by subtracting the mean of image intensities and divide by the standard deviation. New brightness value can be found by using the neighborhood of a pixel in the input image. The fig 2 below shows the preprocessing of image.

- FEATURE EXTRACTION –
  Its main motive is to simplify the image by considering only the information which we want i.e. important information and leaving out the extra information which is not necessary for recognition. It uses the method of edge detection which can only keep the essential information. It represents the reduced part of the required image as a feature vector. This approach is mostly used when the size of image is very large. Hence, by this process, image recognition becomes simplified. It starts from the already measured data and features which provides some kind of information facilitating the further steps.



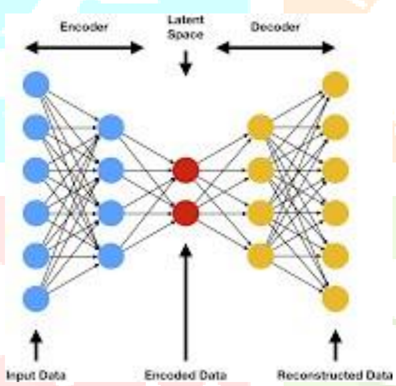
Feature Extraction
Figure no.(17)



Extraction Flow
Figure no.(18)

- DETECTION INIMAGES

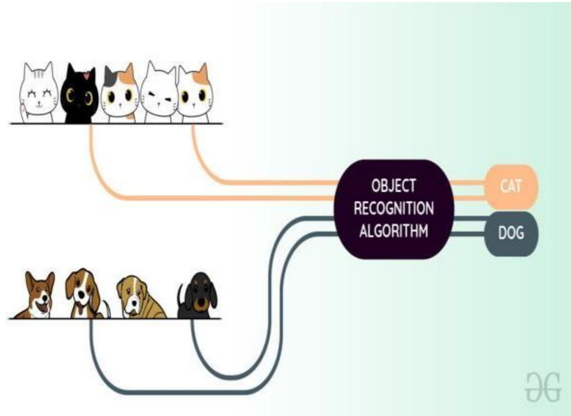
The n number of different components of object detection are integrated into a single neural network, which uses certain features from the whole image to predict a bounding box The bounding boxes for other classes are also predicted at the same time. Hence the neural network analyses the full image and also the different objects in the image.



Detection
Figure no.(19)

- Object Recognition

Object recognition is a computer vision technique for identifying objects in images or videos. Object recognition is a key output of deep learning and machine learning algorithms. When humans look at a photograph or watch a video, we can readily spot people, objects, scenes, and visual details.



Object Recongition
Figure no.(20)

IV. CONCLUSION

Deep Learning based Object Detection has been a research hotspot in recent years. This paper provides a detailed review of deep learning-based object detection frameworks. The need for object detection systems is gaining more importance. Object Detection algorithms typically leverage machine learning or deep learning to produce accurate results. Some algorithms are used to detect object such as OpenCV is the huge open-source library for computer vision, machine learning, and image processing and now it plays a major role in real-time operation which is very important in today's system by using it one can process images and videos to identify objects and instances. YOLO(You Only Look Once) is one of the most effective object detection algorithms Yolo uses totally different approach for object detection process frames at the rate of 45 fps (larger network) to 150 fps(smaller network) which is better than the real-time network is able to generalize the image better. Also, CNN (Convolutional Neural Network)algorithm is used for more accuracy in the detection of objects. It consists of several different layers such as the input layer, at least one hidden layer. They are best used in object detection for recognizing patterns.

By using this thesis and based on experimental results we are able to detect object more precisely and identify the objects individually with the exact location of an object.

Acknowledgment

I would first like to thank our project mentor Prof Geeta Atkar who gave me the golden opportunity to do this wonderful project which Also helped me in doing a lot of research and I came to know about so many new things. I am really thankful to her for providing us with continuous support and guiding us in right whenever we needed it. This research would not have possible without her expert guidance and motivation. Secondly, I would like to thanks my project partners who helped me a lot in finishing this project.

References

1. Pierre Sermanet, David Eigen, Xiang Zhang, Michael Mathieu, Rob Fergus, Yann LeCun, *OverFeat: Integrated Recognition, Localization and Detection using Convolutional Networks*, ICLR 2014 conference submission.
2. Ross Girshick, Jeff Donahue, Trevor Darrell, Jitendra Malik, *Region-based Convolutional Networks for Accurate Object Detection and Segmentation,* IEEE Journals & Magazine.
3. Ross Girshick, Fast R-CNN, 2015 IEEE International Conference on Computer Vision (ICCV)
4. Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi, *You Only Look Once: Unified, Real-Time Object Detection,* IEEE Xplore: 12 December 2016
5. Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, Alexander C. Berg, *SSD: Single Shot MultiBox Detector*, Springer International Publishing. Published in: Computer Vision – ECCV 2016.
6. Shaoqing Ren, Kaiming He, Ross Girshick, Jian Sun, *Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks ,* IEEE Transactions on Pattern Analysis and Machine Intelligence
7. Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, Serge Belongie, *Feature Pyramid Networks for Object Detection,* 2017 IEEE Conference on Computer Vision and Pattern
8. Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, Piotr Dollár, *Focal Loss for Dense Object Detection,* IEEE Xplore.
9. Joseph Redmon, Ali Farhadi, *YOLO9000: Better, Faster, Stronger,* IEEE Conference Publication.
10. Joseph Redmon, Ali Farhadi, *YOLOv3: An Incremental Improvement,* International journal of computer vision
11. Yanjia Li, *Dive Really Deep into YOLO v3: A Beginner's Guide,* yanjia.li
12. Xingyi Zhou, Dequan Wang, Philipp Krähenbühl, *Objects as Points*, IEEE Open Journal of Circuits and Systems
13. Yanjia Li, *Human Pose Estimation with Stacked Hourglass Network and TensorFlow,* yanjia.li
14. Kaiming He, Georgia Gkioxari, Piotr Dollár, Ross Girshick, *Mask R-CNN,* IEEE Conference Publication
15. Zhaowei Cai, Nuno Vasconcelos, *Cascade R-CNN: Delving into High Quality Object Detection,* IEEE Xplore.
16. Mingxing Tan, Ruoming Pang, Quoc V. Le, *EfficientDet: Scalable and Efficient Object Detection*, IEEE Xplore.
17. Jifeng Dai, Yi Li, Kaiming He, Jian Sun, *R-FCN: Object Detection via Region-based Fully Convolutional Networks,* NIPS'16
18. Yanjia Li, *Witnessing the Progression in Semantic Segmentation: DeepLab Series from V1 to V3+,* yanjia.li
19. Golnaz Ghiasi, Tsung-Yi Lin, Ruoming Pang, Quoc **V.** Le, *NAS-FPN: Learning Scalable Feature Pyramid Architecture for Object Detection,* CVPR 2019
20. Shu Liu, Lu Qi, Haifang Qin, Jianping Shi, Jiaya Jia, *Path Aggregation Network for Instance Segmentation,* 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition
21. Yanjia Li, *10 Papers You Should Read to Understand Image Classification in the Deep Learning Era,* Yanjia Li
22. Deep Learning (Adaptive Computation and Machine Learning series) by Ian Goodfellow, Yoshua Bengio, Aaron Courville, Francis Bach.
23. Deep Learning for Natural Language Processing: Applications of Deep Neural Networks to Machine Learning Tasks by Pearson Learn IT.
24. Deep Learning with Python by Francois Chollet.
25. Advanced Deep Learning with Keras by Rowel Atienza.
26. Deep Learning in Object Detection and Recognition.
27. Advanced Applied Deep Learning.
28. Object Detection in Low-spatial-resolution Aerial Imagery Using Convolutional Neural Networks.
29. Hierarchical approach for object detection using shape descriptors.
30. Application of Deep Learning in Object Detection.
31. Static Object Detection in Image Sequences.
32. Moving object detection using GMM on mobile devices.
33. Real Time Video Processing and Object Detection on Mobile.
34. OBJECT LOCALIZATION AND DETECTION USING ANT COLONY OPTIMIZATION.
35. Invisible Nature of an Object Exposed to Radar detection.
36. Novel Approach for detection of objects in surveillance videos.
37. Object Detection with Background Subtraction.
38. Algorithms for detection of objects in image sequences captured from an airborne imaging system.
39. Multi Robot Cooperative System for Object Detection.
40. Object detection Complete Self-Assessment Guide.