Mody University International Journal of Computing and Engineering Research
3 (2), 2019, 90-97
http://www.publishingindia.com/micer

A Review on Object Detection With Deep Learning

Arpit Kumar Sharma[1*], Siddharth Jain[2] and Chirag Goyal[3]

[1]Manipal University Jaipur, Jaipur, Rajasthan, India. Email: er.aks31@gmail.com
[2]Arya Institute of Engineering Technology and Management, Jaipur, Rajasthan, India.
Email: sidjain2910@gmail.com
[3]Arya Institute of Engineering Technology and Management, Jaipur, Rajasthan, India.
Email: chiraggoyal0410@gmail.com
*Corresponding Author

Abstract: Object detection makes a wide placement to make a review on it. Object detection play vital role in getting a proper pictures and video analysis. So due to object detection we will get that proper pictures and videos. Object detection can be done with the deep learning. Deep learning increases the accuracy in object detection. The project has aims to get the perfect object detection with the high level of accuracy. And that accuracy can be achieve by deep learning and show real time performance. We will discuss all the important tool of deep learning. Then we will go through generic object detection and with its types like region proposal generation and classification or regression method. There we will get all methods like R-CNN, fast-R-CNN, faster R-CNN, and tasks which dependent on each other like CNN with SPP. And some more like YOLO and SDD and etc. Each method has their unique property also. We will study about that topic in briefly. On that basis there is challenge of having dependency of object detection on the computer vision system with the help of deep learning. Experiment analysis is also providing to get the different between different types of method of object detection. There is a network which is trained on the most challenging publicly database which is PASCAL VOC, on which object detection is done by annually.

Keywords: Convolutional network, Deep learning, Image, Neural network, Object detection.

I. Introduction

To make an understanding on the object recognizing/detection we should classify different objects and localized the detail of object in an image. There area different types of detection like salient detection, face recognizing, face expression recognizing, pedestrian detection and skeleton detection [1]-[3]. We use the concept of object detection to overcome the problem of fetching detail of an object by a computer system [4]-[5].

II. Problems Statements

Computer faces many of problems to detect an object. In which one of the major problem is the classification of object, here the classification means to predict the class of an object. The one more slightly main problem is to detect the location of image, where image contain no of/single object [6]-[8]. So the main function of a system is to predict the class of location of object, which is like bounding box around the object [9]. So for a system an input will be an image containing object and the output will be bounding box corresponding to all objects in image, and each object of image will be in a bounding box [9]. [9] which is shown in following Fig. 1.

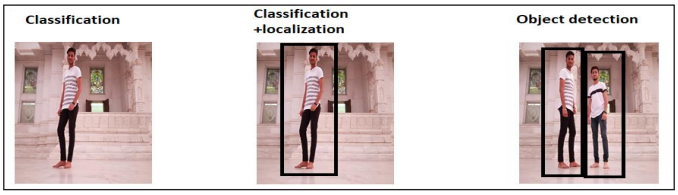

Fig. 1: Object Detection

III. Application

There is some application which is perfectly related to object detection:

- A well known example of application of object detection is capturing a picture from a camera. While capturing a picture through camera, camera will first detect the image and then localized the object by making it in bounding box [10].
- The second more example of object detection is, used in autonomous car. Where sensor of that car will detect the object without conduction of driver.

- That is also used in surveillance system to detect the objects of a particular area.
- These system can be integrated with other task such like pose estimation, here the first stage is to detect the object in pipeline, and the second stage is to estimate the pose of that object in earlier detected region [11].
- Well, by means of object, this is also used to detect the object, through which they can be used in medical and robotics application.
- There are some more application of object detection like face recognizing and face expression analyzing.

IV. Concept of Deep Learning in Object Detection

We can enhance the techniques of object detection by using deep learning. It uses the bounding box to bind an object in an image. In deep learning neural network algorithm [12]. We will define deep learning in DNN (deep neural networks) and CNN (convolutional neural network) and YOLO (you only look once) algorithm [12]-[14]. Which is shown in following Fig. 2.

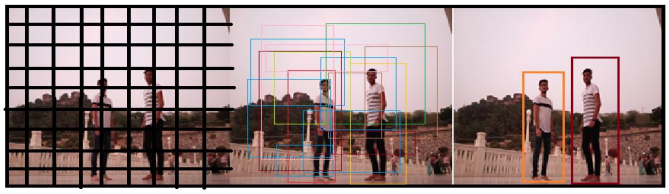

Fig. 2: YOLO Algorithm for Object Detecting

V. Challenges

The major challenge of that problem is that of the dimension of the detect object in our output that are caused due to variable number of objects in a given input image. This will be challenge because a general machine learning task requires a fixed dimension of input and output for the model to be trained. The one more challenge is that widespread adoption of object detection for a real time. Where the obstacle is to fine them accurate. The more complex model, it require more time to interface where for less complex model is less in accuracy. So there is clash between accuracy and performance, there is need of choosing as per our application.

A. Terminology of Object Detection

There are followings of terminology of object detection:

- *Object Localization*: It is the process of determining the actual position of an object at which the object is located in an image [15].
- *Object Classification*: It is the process of categorized an object from different objects in an image [16].

B. Methodology for Object Detecting

1. Model of Object Detection

There are three model of detection which are following:

- *Informative Region Selection*

It is a process of differentiate an object from a set of object with different size and dimension in a image it a type of neural choice to detect the image or set of object with multi-scale sliding windows. If there is a lot of sliding windows then there is chances to get many redundant windows [16]. Beside it if we take less number of windows then there is chances to get unsatisfied result.

Bounding Box

The bounding box is rectangle box which will drown on image where the object inside the image will totally fit in that box. A bounding is made for all object instances in the image. The distance measure a jacquard distance which computes intersection over union between the predicted and ground truth boxes shown in Fig. 3.

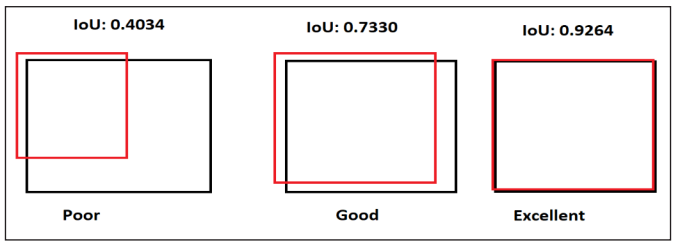

Fig. 3: Jacquard Distance

- *Feature Extraction*

In this we determine the object with its different features which provides a important knowledge, information and visual representation. There is difficulty to design a descriptor which determine the all types of object because of occurrence of different type of object appearance, lighting condition and background/ color effect [16]-[17].

- *Classification*

In the classification method we classify and categorized object and details of object to make a visualize representation. There are some choices like DPM, SBM, ADBOOST. In this DPM is more reliable for object detection [17]. In DPM we design low level algorithms with the help of graphical model that is usable for high precision part model for a variety of object.

Object detection is based on the local feature detection which can be obtained on the PASCALVOC.

Deep learning means DNN and more reliable region CNN(R-CNN). It has an architecture which is sued to detect and fetch detail of an object (local and shallow part detection). It has algorithms which allow to learn information without detecting it manually like YOLO algorithms.

2. Two Stage Method

In this case, proposal are extracted using different computer vision technologies and realized to fixed the input for the classification network, which acts as a feature extractor. Then a SVM is used to classify the object and background (one SVM for each class). Also a building a box regressor is trained that output some correction (offsets) for proposal boxes. These method is very accurate but are computationally intensive.

3. Unified Method

The difference here is that, instead of producing proposal predefine a set of object boxes are already made and just look for objects. Over his action of maps the map will predict class scores and building boxes [18].

There are some steps to follow this method which are following:

- Build a CNN with regression and classification objective.
- Gather all the activation from later layer to infer classification, and location with a fully connected or convolutional layer.
- During building, use jacquard distance to predict the distance of ground truth.
- During interface, there will a such condition of having multiple boxes on a single object is occur at that condition non maxima method id used to remove that thing.

4. CNN (Convolutional Neural Network)

It is most important model of deep learning. VGG16 play a vital role in CNN as an architecture. There are different types of layer in neural network as input layer, at least one hidden layer and a output layer which are used in object detection. They recognized edges (horizontal or vertical), shapes, colors, background and textures. In this the hidden layer uses as convolutional layer, which works like a filter. It firstly get an input then convert into a specific pattern/feature and send to next layer. There are a no. of filter i.e. convolutional layer. Every time a new input is sent to the next level of layer. That can be understand with an example - in the first convolutional layer we filter or identify shape/color of an object (i.e. black) then next one may be able to conclude the object in a region (i.e. paw, beak and wings). And at last convolutional layer may classify the object as a pigeon bird. There may be more intermediate layers between these layer that have input and transmit to next layer.

But R-CNN is a slow process so we will move to next process.

5. Generic Object Detection

Generic object detection have aims to classifying and locating objects in a image. The objects are localized with making object in a rectangle bounding boxes. Generic object detection have two main method to detect an object [18]. And the other one is taken like classification and regression problems. Where objects are firstly classified and then localized. Region proposal generation include mainly R-CNN, SP-net, fast R-CNN, aster R-CNN and some of them which are corelate to each other like CNN with SPP. And classification or regression method based on mainly YOLO and SSD.

VI. Region Proposal Generation

Here we will discuss about R-CNN, SPP-net, Fast R-CNN and Faster R-CNN.

A. R-CNN

R-CNN is a region proposed base process of object detection that is proposed by Ross Girshick in 2014. R-CNN have a search to provide almost 2000 region proposal for each image. It is mainly three way process:

- It is a process of detect an image for possible object. We use region proposal algorithms for search, generating 2000 region proposal.
- We run a CNN on the top of region proposal.
- In this we use SPM to categorized a linear regression to make a boundary box of an object more tightly (i.e. exactly).

This is shown in following Fig. 4.

There are so many application like autonomous driving, smart surveillance system, facial reorganization.

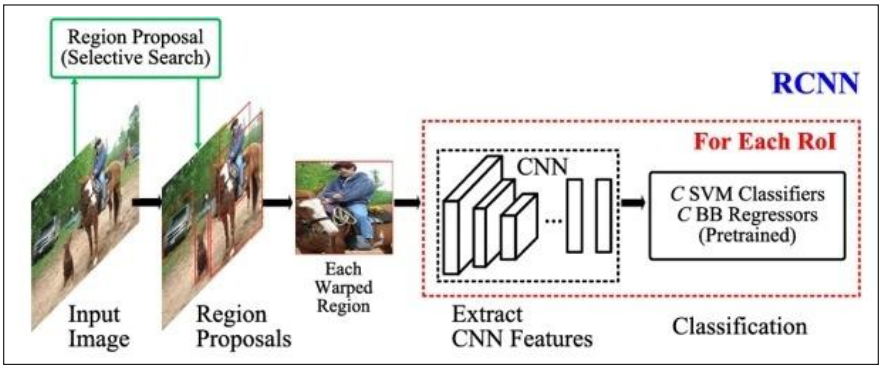

Fig. 4: RCNN

B. SPP-net

We know that CNN require a fixed size of image. That fixed size is 224*224. Then we adopt another way of pooling is called spatial pyramid pooling is used to solve the above requirement. The theory of special pyramid matching provide a scale to partition the image into a number of division and aggregates of local feature into mid level representation which is shown in following Fig. 5. The layer after the CONV-5 act as SPP layer. It use 256 features match in CONV-5 with three level pyramid then final obtain result proposed by SPP has a dimension like.

256*(1^2+2^2+4^2) = 5376

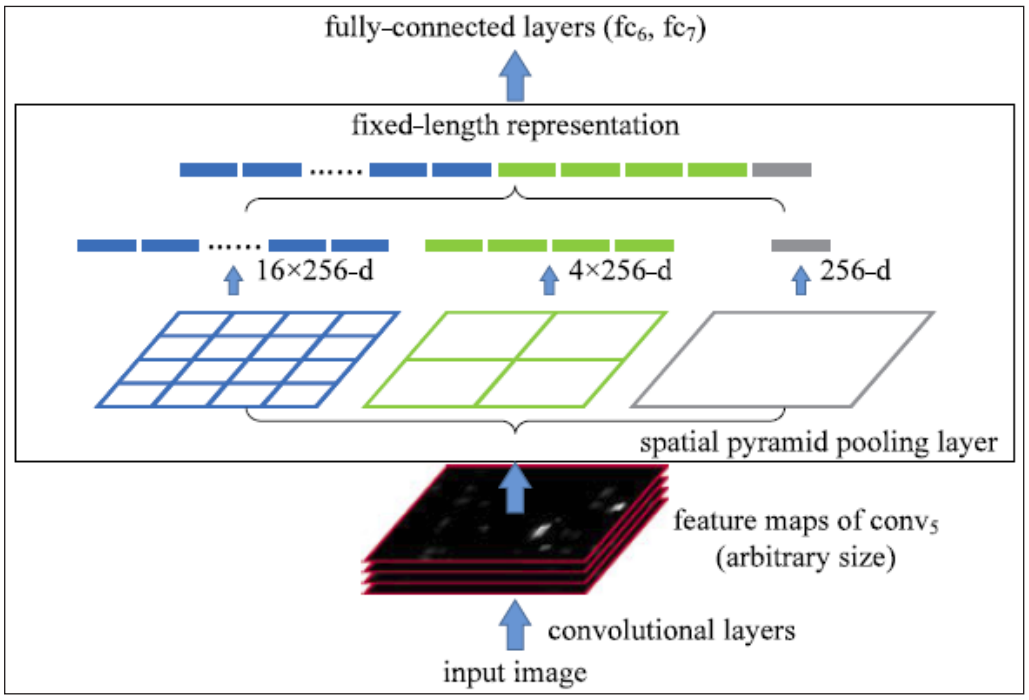

Fig. 5: SPP-net

C. Fast R-CNN

We use CONV layer in SPP but it is enable with five turning algorithm then an accuracy is dropped with R-CNN. So Ross Grishick proposed in novel convolutional network architecture is called Fast R-CNN.

n the Fast R-CNN an output layer has main role to produce SoftMax probabilities for all C+1 categorized and other output layer works as to encode bounding box position with four real named numbers which is shown in following Fig. 6.

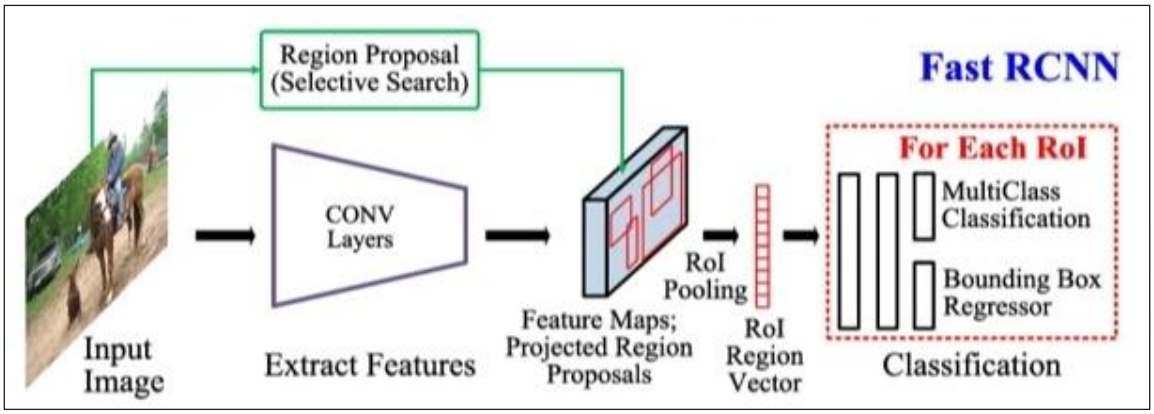

Fig. 6: Fast R-CNN

D. Faster R-CNN

It is also proposed by Ross Girshick and it is most famous architecture of object detection it uses CNN as YOLO and SSD algorithm. It is 10 times faster than Fast R-CNN, even they have same accuracy of data set. It replaces the selective search method IRPN which is much faster which is shown in following Fig. 7.

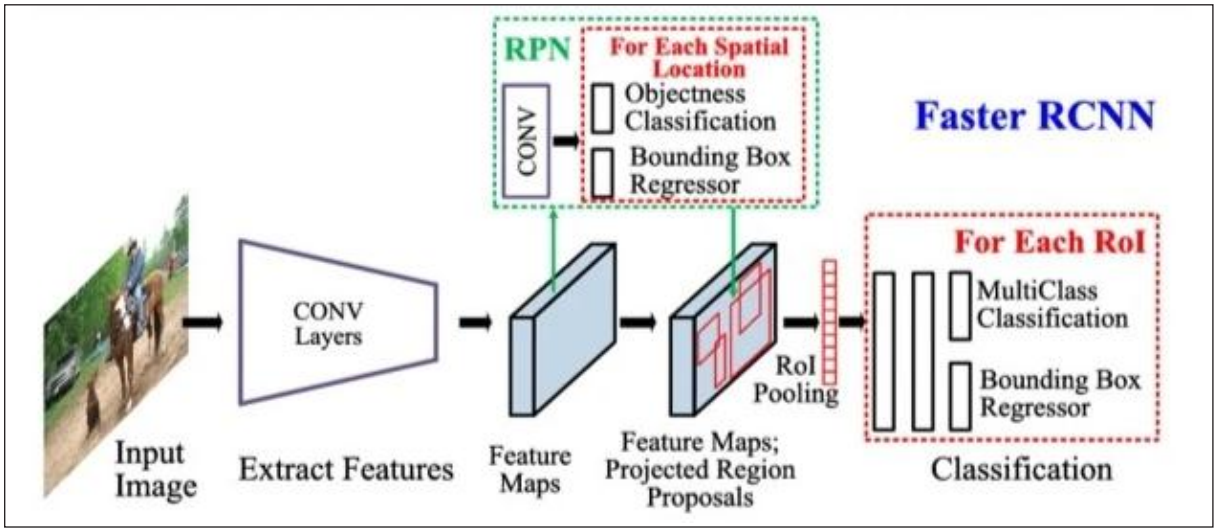

Fig. 7: Faster R-CNN

VII. Classification or Regression Based Method

There are so many stages in region proposal based framework like region proposal generation, feature extraction with CNN, classification and bounding box regression but here we will discuss on two main significant framework which are based on regression/classification. That are YOLO (you only look once) and SDD (single shot multibox detector).

We will discuss about YOLO and SSD in briefly as following:

YOLO

YOLO is unified detector casting object detection that is proposed by YOLO Redmon as a regression problem. The problem is related to image pixel, bounding box which separate an object and so many related class probabilities which is shown in following Fig. 8. .

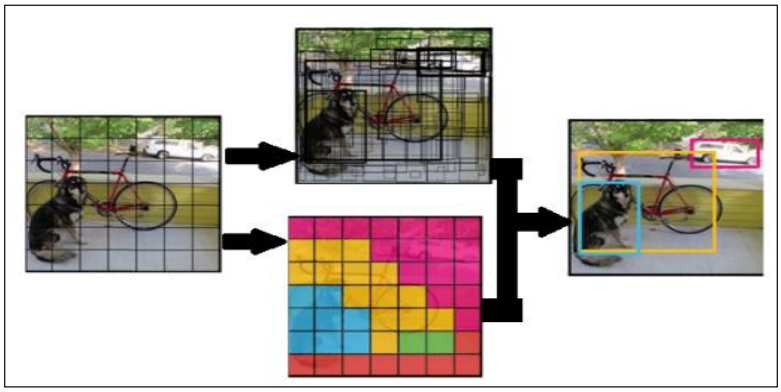

Fig. 8: YOLO

YOLO use a tiny set of candidate region to detect an object.

The main mechanism of YOLO is –

First it divide an image into a S*S grid, each detecting C class probabilities, B bounding box locations and confidence scores. Instead of RPG (Region Proposal Generation) step, YOLO is more fast by design, its real run time is 45FPS and fast run time is 155FPS.

YOLO fails to localized an object than fast RCNN because of the presence of only one object in grid, coarse grid division. So there are two expand version of YOLO in which custom GOOGLENET network is replaced by simple DARKNETS, normalization, anchor boxes that are YOLOV2 and YOLO09000. They provide an approach of dataset and COCO detection with word tree to combine data.

SSD

SSD means single shot MultiBox detector. It is used to provide real time speed without detecting accuracy. It is proposed by Liv, that is more faster than YOLO. The main idea of SSD is to provide final detection. SSD used multiple CONV feature map to perform the task of detection which is shown in Following Fig. 9.

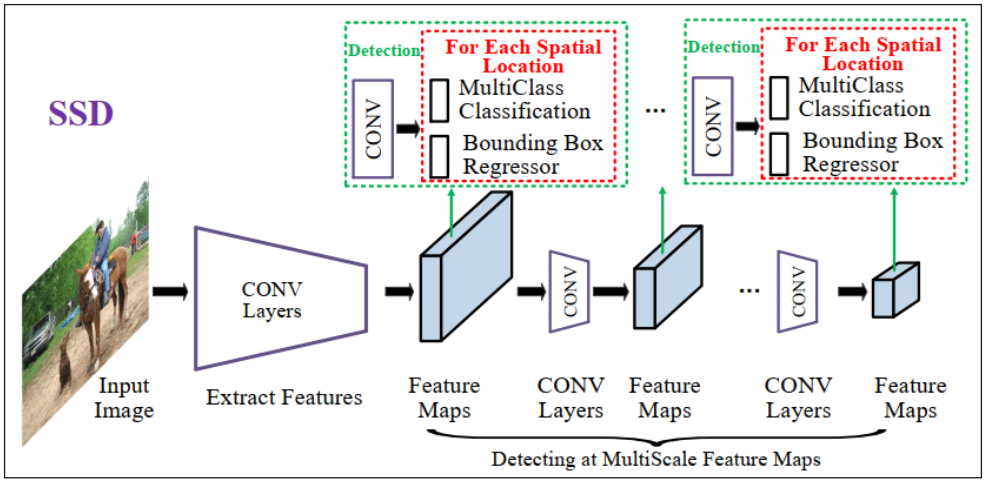

Fig. 9: SDD

VIII. Some Task That are Reach by Object Detection

There are many task which are done by using object detection. From which here the followings are some of types of that tasks:

A. Salient Detection

Salient detection is the one of the most important task, which is mostly used to detect the most dominant object region in an image. There is a incorporation of numerous application in visual saliency to improve their performance such like image cropping, segmentation, image retrieval and object detection.

Broadly there are two types of salient detection where one is bottom up (BU), and second one is top down (TD). To know the information about local contrast, various local and global things are come to know e.g. edges and spatial information. Local feature is low level feature that can not explored the high level and multi scale semantic information. So low level maps are obtain at the place of salient object detection.

Due to these all application CNN must have to use in it for more accuracy. That shoe the need of deep learning in salient detection.

The process of salient detection using deep learning can be clear by following Fig. 10:

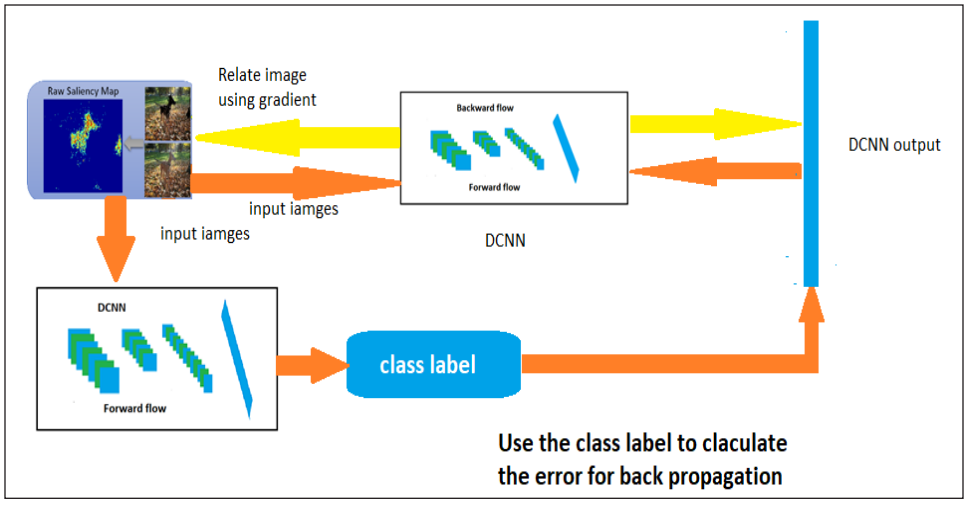

Fig. 10: Salient Detection Using Deep Learning

B. Face Detection

Face detection is computer based application which is used to identify the human face in digital form of image. They can be overcome by using of deep learning process.

Face detection is generally refers to a frontal human or person and by means face detection is only the technique which is used to detect the face of person who are exact front of system,

This is the process of matching the image of that person which is captured by system is matched by bit. The image is matched by the storing images from bit to bit. The stored image are stored in database. If there is some change in database image, then the match will be unmatched.

The main moto of face detection is to analyses and recognize the face expression. It covers a large range of scales (30-300pts vs 10-1000pts). Each face have unique object structure like different face parts and characteristics like skin color. So it is challenge to detect different type of face which have variation of pose, expression, color, conclusion.

Deep Learning in Face Detection

Flezenszwalb proposed a cascade structure which is based on deformable part model (DPM) for face detection. YU proposed a novel LOV function which is based on CNN to detect the four bounds of box jointly which is shown in following Fig. 11.

YANG provide two networks:

First one is based on novel deep learning based face detection and second one is skill friendly detection network. First one is used to collect the data of facial parts like eye, nose and mouth etc. to address face.

In second one is used to split a large range of target scale into smaller sub range. Huang proposed a unified end to end FCN framework. Which is based on 3D modeling and face land marks called dense box.

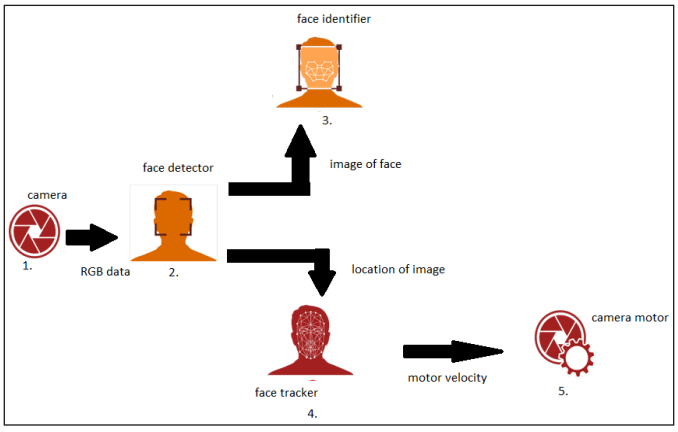

Fig. 11: Face detector

Mainly the face detection are used in cameras and security also. In security purpose there is lock is present where the face id can be inserted. So in those when we add some face id, then they will become the key for that lock. And that whole process is done with the help of object face detector. So when that face come near to the sensor of that lock, that lock will automatically unlocked.

C. Pedestrian Detection

It is an intelligent video surveillance system to detect an essential and significance task. It use to get fundamental information for understand the all the activity like pedestrian tracking, person identification. There are many types of size in pedestrian detection e.g. automatic driving and an intelligent surveillance. We have the application of ROI pooling layer in generic object detection.

Pedestrian detection may be failed if there is a default in background and hard light effect and blurredness of objects which is shown in Fig. 12.

Role of Deep Learning

DCNN play a vital role in pedestrian detection.

Zhang adopt a generic faster R-CNN method to perform pedestrian detection. We use the concept of CONV feature maps to modify pedestrian detection.

Tian proposed a deep learning framework to deal with complex occlusion existing in pedestrian image called Deep parts.

We use CNN base model because it is most prior method that provide more accurate candidate boxes and multilevel information to recognized and localized the pedestrian.

To achieve better result, we combine CNN with the handcrafted feature that are complimentary. So we can do better pedestrian detection.

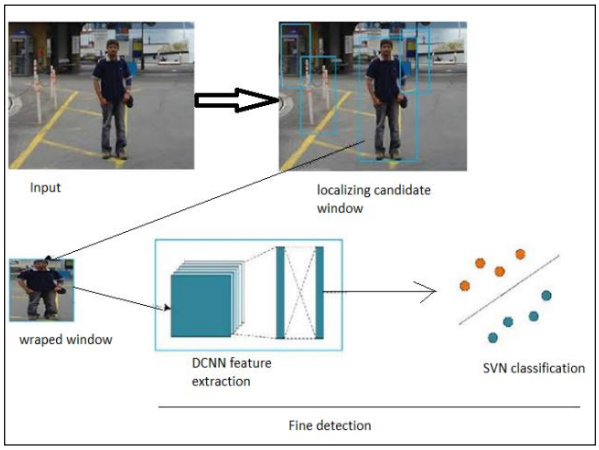

Fig. 12: Pedestrian detection

IX. Conclusion

Object detection become the most challenging task. This is because the number of object are varies from an image to another

image. That make the challenging condition. There is two stage for object detection the first one is to detect the object directly. And the second one is to perform region proposal and then determine the objects in fit bounding boxes. So there is proper dependency of object detection on deep learning. There are we discussed more technologies to get the detection of object like R-CNN, CNN, SP-net, Fast R-CNN, Faster R-CNN etc. here we had discuss more of applications which are more usable at their own base like the face detection which are widely used in cameras, where the whole process of getting accurate picture is based on face detector. Where the salient detector and pedestrian detector with their own base. Where in object detection there is three step to complete the whole process of object detection which are: classification, localization + classification and detection. Where in image consider in classification and object fit in bounding box in localization + classification and in detector we get proper detection of an object of an image. So we can say that the object detection have a great role in real world time and deep learning have co-relation with object detection which show dependency of object detection process on deep learning.

References

[1] X. Wang, M. Yang, S. Zhu, and Y. Lin, "Regionlets for generic object detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 10, pp. 2071-2084, October 2015.

[2] K. Gautam, V. K. Jain, and S. S. Verma, "A review on vehicular communication system," *A Journal of Composition Theory*, vol. 12, no. 9, pp. 2037-2041, 2019.

[3] W. Cao, J. Yuan, Z. He, Z. Zhang, and Z. He, "Fast deep neural networks with knowledge guided training and predicted regions of interests for real-time video object detection," *IEEE Access*, vol. 6, pp. 8990-8999, 2018.

[4] P. Prashanth, K. S. Vivek, D. R. Reddy, and K. Aruna, "Book detection using deep learning," *3rd International Conference on Computing Methodologies and Communication (ICCMC)*, pp. 1167-1169, 2019.

[5] J. Wang, S. Jiang, W. Song, and Y. Yang, "A comparative study of small object detection algorithms," *2019 Chinese Control Conference (CCC)*, pp. 8507-8512, 2019.

[6] F. Çakmak, E. Uslu,…and S. Yavuz, "Deformable part model and deep learning comparison on victim detection," *24th Signal Processing and Communication Application Conference (SIU)*, Zonguldak, pp. 1513-1516, 2016.

[7] X. Zhou, W. Gong, W. Fu, and F. Du, "Application of deep learning in object detection," *IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS)*, Wuhan, pp. 631-634, 2017.

[8] P. F. Felzenszwalb, R. B. Girshick, D. Mcallester, and D. Ramanan, "Object detection with discriminatively trained part-based models," *IEEE Transctions Pattern Analysis and Machine Intelligence*, vol. 32, no. 9, p. 1627, 2010.

[9] X. Xu, and E. L. Miller, "Adaptive difference of Gaussians to improve subsurface object detection using GPR imagery," *Proceedings International Conference on Image Processing,* Rochester, September 2002.

[10] U. Anitha, and S. Malarkkan, "Review on sonar image enhancement and object detection using image fusion techniques," *International Conference on Green Computing, Communication and Conservation of Energy (ICGCE)*, pp. 250-253, 2013.

[11] Y. Li, J. Wang, X. Liu, N. Xian, and C. Xie, "DIM moving target detection using spatio-temporal anomaly detection for hyperspectral image sequences," *IGARSS 2018 - IEEE International Geoscience and Remote Sensing Symposium*, pp. 7086-7089, 2018.

[12] L. Lu, J. Geng, and T. Zhao, "Research and analysis small infrared object detection track algorithm and its image processing technology," *3rd International Conference on Computer Research and Development*, pp. 30-33, 2011.

[13] L. Wang, and Y. Liu, "Moving object detection and extraction in serial image," *4th International Congress on Image and Signal Processing*, pp. 928-931, 2011.

[14] A. S. Mohan, and R. Resmi, "Video image processing for moving object detection and segmentation using background subtraction," *First International Conference on Computational Systems and Communications (ICCSC)*, pp. 288-292, 2014.

[15] X. Zhang, H. Li, and L. Jiao, "A change detection algorithm based on object feature for SAR image," *2nd Asian-Pacific Conference on Synthetic Aperture Radar*, pp. 693-696, 2009.

[16] S. H. Kim, N. K. Kim, S. C. Ahn, and H. G. Kim, "Object oriented face detection using range and color information," *Proceedings Third IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 76-81, 1998.

[17] N. Bhargava, A. K. Sharma, A. Kumar, and P. S. Rathoe, "An adaptive method for edge preserving denoising," *2nd International Conference on Communication and Electronics Systems (ICCES)*, pp. 600-604, 2017.

[18] K. Gautam, V. K. Jain, and S. S. Verma, "A survey and analysis of clustered vehicular communication emergency system (CVCES)," *In Press, IEEE Conference*, 2020.