Study On Object Detection Using Computer Vision

MAHIMA TEWATIA[1], DR DEEPIKA PUNJ[2]
[1, 2] *J.C Bose University of Science, and Technology, YMCA, Faridabad, India*

Abstract— Many applications in computer vision they need precise and efficient detection systems. This demand coincides with the rise of the application of deep learning techniques in almost all areas of machinelearning and artificial vision. This work presents a study that encompasses different detection systemsbased on deep learning, providing a unifiedcomparison between different frameworks in order to carry out a technical comparison of the performance measures of the studied methods.

I. INTRODUCTION

Object detection is one of the most researched computer vision tasks, where Convolutional Neural Networks (CNNs) are currently showing very high performance. CNNs are built from many layers of intrinsically connected neurons in a model inspired by the hierarchical organization of the human cerebral cortex. Neurons act as a basic unit in learning and extracting features from input. Theperformance of the learning and the extraction of characteristics of the input is improved with the increase of the complexity of the networks which ismainly caused by the depth of the neuron layers. Deep learning techniques or Deep Learning (DL) ingeneral, and particularly CNNs, They are capable of automatically learning, from generic input images, data with multiple levels of abstraction due to the deep architecture that facilitates the model's capture and generalization process of the filtering mechanism by performing convolution operations in the image domain. In the literature you can find many high performance CNNs such as AlexNet [1], VGG [1], GoogleNet [2], ResNet [3], etc. Some like[2], have been shown to exceed the precision of the human eye in certain object recognition tasks.

Despite the popularity of other methods, DL-based methods are outperforming other traditional computer vision techniques by a wide margin in terms of precision and sometimes even efficiency [4]. However, the changing state of the DL producedby a lack of unifying works and reviews of the state of the art, make Getting started in this field tedious and difficult to keep up to date. Consequently, the below are the elaboration about frameworks which is regressively used for object detection.

A. Tensor flow

Tensorflow is an open source library developed for high performance in numerical computation. It has a flexible architecture and allows easy development fast computing for platforms such as CPUs, GPUs or TPUs. And it's easily scalable which allows it to function from computers to server clusters. It originates from the developers, researchers, and engineers at Google Brain with Google's artificial intelligence organization. This provides great support Strong for Machine Learning and Deep Learning, although the flexibility that this brings Computing is used in many other scientific fields [6].

B. Pytorch

Pytorch is an open source library based on the Torch architecture, being thus a scientific computing package based on python, with special emphasis onthe use of GPU. Since this library was launched in 2016 by Facebook, many researchers have opted to use it because it provides this ability to perform complex calculations of tensors with great GPU acceleration support and ease of cementing deep neural network of enormous complexity [7].

C. Darknet-YOLO

Darknet is a framework developed in C ++ language oriented to the design, training and execution of deep neural networks destined to the detection and classification of objects in 2D images [8]. The main advantages of this system are its simplicity in termsof use, small size, ease of compilation, and clear and concise online documentation. All of them make Darknet-YOLO an easy system to use as soon as it is installed. Also noteworthy is its ability to use NVIDIA's CUDA framework, which allows the system to use the computing power of the GPUs to carry out both training and validation processes.

Included in the Darknet framework, YOLOv2 [9] isa system Detection of objects that applies a single neural network to the entire image using a single evaluation of the network to divide it into regions, which will be used to predict the locations of the bounding boxes, in contrast to other systems such as the Faster-RCNN , whose mode of operation consists of applying aknown CNN model to thousands of regions in the same image. This difference makes YOLO a much faster and computationally lightweight system than other frameworks such as Tensor flow or Pytorch, while maintaining acceptable accuracy rates in its predictions.

When it comes to detecting and locating objects based on CNNs, Darknet is an interesting system because it uses a different algorithm from the classic Faster-RCNN model but achieves similar resultswith a lower computational load, which gives it a higher speed. For this reason, in the realization of this work, two different models of neural networks have been considered to be used with this system: One is YOLOV2 which was the most advancedYOLO model developed by its authors, J. Redmon and A. Farhadi, at the time of the present study. The another is an adaptation of YOLOv2 based on the CNN VGG-16 model that has been carried out in this same work in order to be able to make a comparisonwith the others based on CNN VGG-16 [8-10].

II. LITERATURE REVIEW

A. Computer Vision

Computer Vision is an engineering discipline whichis used to obtain Useful information from captured images. The way to do it is by through the use of different techniques to acquire, process, analyze and understand the information that is in an image. [11-13]. One of the most used attributes in computer vision systems is the detection of objects. the detection by means of characteristics of objects and,more recently, the use of neural networks or systemsof learning. Of the three methods mentioned above,the background removal method is the only one thatdoes not require prior training to determine which objects should detect. Next, we will see an overviewof the operation of each of the methods, as well as the advantages and disadvantages that each of thesepresent.

B. Object Detection by Background Elimination.

This method of object detection is one of the most famous due to the simplicity in which works and because this method is able to detect objects moving precisely. The operation of this method is to provideor generate which is the background of the image to detect, for example, if the camera to be used is static,it is possible to obtain an image of the captured scene that does not contain any of the objects to be detected, otherwise the program must be able to generate and identify which is the background of theimage of the objects that move through it. However,this method presents several problems, such as that the program is not able to differentiate between the detected objects, in addition to changes in light and the slightest change on the position of the image may affect the way of detection. [14] [15] [16] differentiate between detected objects, as well as changes in light and the slightest change on the image position may affect the way of detection. [14][15] [16].

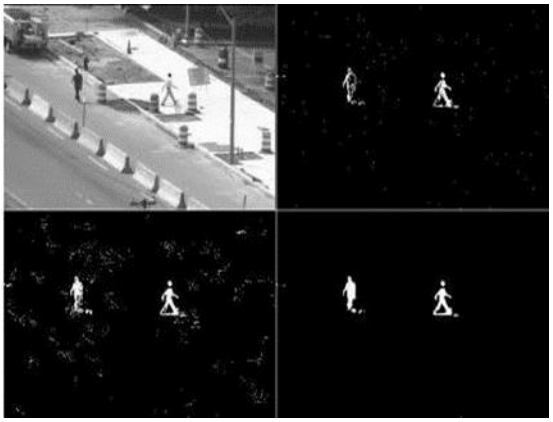

Fig. 1: Example of object detection by background removal

In fig 1, the original image is seen in the upper left before going through the object detection algorithm.In the upper right image the result is shown after comparing the original image with the previously defined image as the background of the image, in this it can be seen that the detector has found several subjects, however one of them is not completely defined, the result of this is the lower right image in which you can see a more defined image of the objects detected and with a significantly lower amount of noise than in the previous images.

C. Object Detection by means of HAAR Classifiers.

The detection of objects by means of HAAR classifiers consists mainly in the use of recognition cascades of the HAAR characteristics to learn to detect the edges and corners of a specific type of object. The HAAR characteristics are obtained by using a pair of adjacent rectangular geometric shapes of the same size to obtain the difference in the contrast between both, this difference is obtained by adding the contrast of each of the pixels contained in each of the rectangles and then simply the difference between the two is obtained. This process is performed on the entire image by moving the used rectangles pixel by pixel and re-analyzing the image by changing the size and shape of the used rectangles, in this way it is possible to detect objectsof different sizes and shapes. Finally a filter is applied which limits the minimum value to eliminate the possibility of obtaining noise. [17] [18] [19]

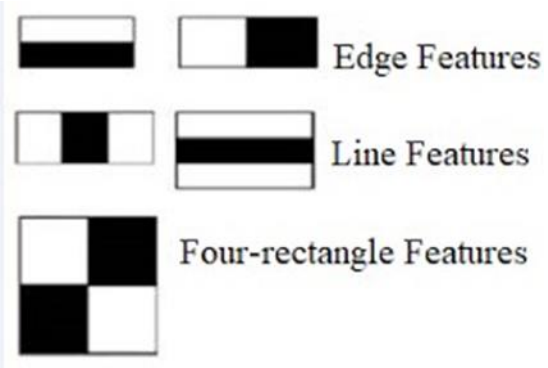

Fig. 2: Examples of HAAR Characteristics

This type of object detector has the advantage that the detection becomes more accurate, in compared to background removal method, because the HAAR method will only identify those objects that you want to detect.

The reason for this is that a previous training using hundreds or thousands of images containing only the objects that necessary, which significantly reduces the number of objects that are detected at the time torun the algorithm.

However, one of the disadvantages of using this method is in the fact that the use of this algorithm is susceptible to changes in the illumination of the image to be analyzed, since this can cause false detections or that it is incapable of detect objects. This is mainly due to various factors related to the images used during training (quality of the image used, contrast, luminosity, position of the object, among others) establish which are the patterns that define the most prominent characteristics of each ofthe object classes and, once obtained these class patterns, use the patterns obtained during training tobe able to Identify in any other image to see if thereis a match between these and the image. [17] [18].

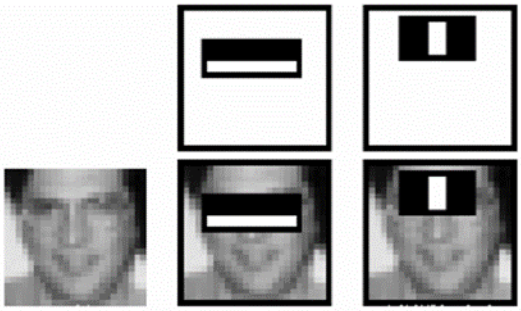

Fig. 3: Recognizing HAAR Features in an Image

D. Object Detection using YOLO (You Only Look Once)

YOLO is an object detection and classification method which involves the use of convolutional neural networks, or CNN for its acronym in English (Convolutional Neuronal Network), to perform the detection. Compared to other types of methods which use convolutional neural networks, YOLO only requires a single layer neural network, instead of using several layers as in the case of the RCNN (Region with Convolutional Neural Network) system and its derivatives, which makes the operation of this much faster. The operation of YOLO is as follows, each frame of the video is divided into a grid of size S x S, where each of the divided regions is subjected to an analysis to the neural network, where each neuron of the single layer corresponds to each member of the classes in which the system has been previously trained, this in order to know to which class each detected object belongs in each of the cells, how exact the detected classification is and at the same time determines thedetection box of each of the objects. [19] [20] [21] [22] [23] [24].

The advantages of this detection method are the speed and precision in the detection and classification of the objects to be detected as long asthey are within the classes in which the program hasbeen trained, this is

achieved since the system givesan accuracy value to each detected object which corresponds to the classes with which it finds similarities, so it is possible to reduce the number ofdetections incorrect when setting the desired accuracy limit value.

The disadvantages of this method can be several, one of which involves the training part, since if a training is carried out with a poor image set, the detection is likely to fail, in addition to this, if the number of classes used in the training is very long, this can cause the system to take longer to perform the detection. On the other hand, as the system useseach of the regions in which it is initially divided toperform the detection and the creation of the detection box, if the size of the object is too small.

III. LITERATURE STUDY

Researcher	Title	Purpose	Conclusion	Suggestions
R. Ranjan et al. (2017) [26]	An All-In-One Convolutional Neural Network forFace Analysis	For face detection simultaneous, face alignment, pose estimation, gender Recognition, detection smile, estimated age, and facial recognition	MTL method framework much better than hyperfaces. MTL helps learning process feature descriptors	In the future, This research is planning to extend method fortax and make real algorithm time.
Cordeiro et al. (2018) [29]	A Convolutional Neural Network with Feature Fusion for Real-Time Hand Posture Recognition	Convolutional Neural Network (CNN) aims to posture recognition hand	In part large dataset evaluated, pictures binary can give rate recognized equivalent content with depth or grayscale representation.	Researchers suggest to investigate another method for selection hyperparameter and optimization, such as multi-objective algorithm.
Guo et al. (2020) [30]	A Fast Face Detection Methodvia ConvolutionalNeural Network	Aims to find method fastest for face detection with convolutional neural network	Which method used for direct classification on dcfs, by significant can increase efficiency during Detection Process face	propose method Fast face detection based on Extracted DCFs by CNN
Ding et al. (2018) [31]	Trunk-Branch Ensemble Convolutional Neural Networks forVideo-Based Face Recognition	To learn representation of the face blur on video with Convolutional Neural Networks (CNN)	The TBECNN approach proposed by effective able handle pictures blurring, occlusion, and pose variations	Suggestions for using the framework that comprehensive based on Convolutional Neural Networks (CNN) and CNN architecture named TBE-CNN.

Ahsan, Mominul & Based, Md & Haider, Julfikar & Kowalski, Marcin. (2021) [32]	COVID-19 Detection from Chest X-ray ImagesUsing Feature Fusion and Deep Learning.	Detect the COVID-19 Symptoms using Lungs or Chest X-Rays	Histogram-oriented gradient (HOG) and Convolutional neural network (CNN) from X-ray images were fusedto develop the classificationmodel through training by CNN (VGGNet).	Further Classes Like Pneumonia, Edema, Hernia and Fibrosis are not classified therefore ResNet and Alexnet as combined model can produce the results of missing classes.
S. Iqbal et al(2021) [33]	Prostate Cancer Detection Using Deep Learning and Traditional Techniques	Detecting Cancer using Deep Learning Models like CNN and LSTM	The results show that ResNet - 101 deep learning outperformed than non-deep learning methods and LSTM. Thus, the deep learning method	Faster RCNN and LSTM can compilethe model more swiftly with less computational resources attaining the high accuracy.

Researcher	Title	Purpose	Conclusion	Suggestions
Zhang et al, (2012) [25]	Automatic Facial Expression Recognition Basedon Hybrid Features	For facial expression recognition from the front face image display automatically	With use endorsement machine classifier vector (SVM), which method can be proposed reach level confession 87.33%.	In automatic facial expression recognition: feature localization and feature extraction, an adaptive initialization method for the AAM model is proposed, which is more suitable for the feature point localization of facial images of expression variations. And a hybrid feature set, consisting of AAM features of shape, geometry, and appearance.
Jalali et al. (2017) [27]	Sensitive Deep Convolutional Neural Network for Face Recognition at Large Standoffs with Small Dataset	Get additional sensitivity in the training structure Convolutional Neural to highlight frequency component small height and variation	Characteristics that created from that model proposed more can be distinguished leads to common exploitation, internal and face recognitionbetter	Using blurred bimodal for faces recognition with longdifferent make more algorithms sensitive to change Intensity.

		from the sample image (dataset)		
Trigueros et al. (2017) [28]	Enhancing Convolutional Neural Networks for Face Recognition with Occlusion Maps and Batch Triplet Loss	To know which part of the face more human important to achieve high level of accuracy high with help convolutional neural network	The best CNN models generated from combination two approaches proposed from this research	Proposed occlusion map for visualize grafting area, on the same time and classification modelon that face showing occlusion artificial.
Cordeiro et al. (2018) [29]	A Convolutional Neural Network with Feature Fusion for Real-Time Hand Posture Recognition	Convolutional Neural Network (CNN) aims to posture recognition hand	In part large dataset evaluated, picturesbinary can give rate recognized equivalent content with depth or grayscale representation.	Researchers suggestto investigate another method for selection hyperparameter and optimization,such as multi-objective algorithm.

IV. CONCLUSION

Theoretical considering the existence of various types of neural networks and their use in various professional fields, the systematization and consolidation of theoretical information has been carried out from the necessary approach to be able to design a predictive model for the early detection of various objects and entities using Deep Learningand Computational Vision. Currently there are many investigations for the detection of various objects using neural networks, the present work seeks to obtain the best theoretical / practical knowledge of works already carried out, to obtain the best results and be the basis for future research that can expand the general purpose. Likewise, it is considered that the analyzes carried out based on the images obtained by international and national data sets offer a higher precision index, in order to obtain better results in the early detection of various objects using Computer Vision.

REFERENCES

[1] Verdhan, Vaibhav. (2021). VGGNet and AlexNet Networks. 10.1007/978-1-4842-6616-8_4.

[2] Gogineni, Saikiran & Pimpalshende, Anjusha & Goddumarri, Suryanarayana. (2021). Eye Disease Detection Using YOLO and Ensembled GoogleNet. 10.1007/978-981-15-5258-8_44.

[3] Amaya Rodríguez, Isabel & Civit-Masot, Javier & Luna- Perejon, Francisco & Duran-Lopez, Lourdes & Rakhlin, Alexander & Nikolenko, Sergey & Kondo, Satoshi & Laiz, Pablo & Vitrià, Jordi & Seguí, Santi & Brandao, Patrick. (2021). ResNet. 10.1007/978-3-030-64340-9_12.

[4] Ahmad, Jamil & Farman, Haleem & Jan, Zahoor. (2019). Deep Learning Methods and Applications. 10.1007/978-981-13-3459-7_3.

[5] Chauhan, Rahul & Ghanshala, Kamal & Joshi, R.. (2018). Convolutional Neural Network (CNN) for ImageDetection and Recognition. 278-282. 10.1109/ICSCCC.2018.8703316.

[6] Ekman, M. (2021). Learning Deep Learning: Theory and Practice of Neural Networks, Computer Vision, Natural Language Processing, and Transformers Using TensorFlow (1st ed.). Addison-Wesley Professional.

[7] Papa, J. (2021). PyTorch Pocket Reference: Building and Deploying Deep Learning Models (1st ed.). O'Reilly Media.

[8] Bharat Sikka, Elements of Deep Learning for

Computer Vision: Explore Deep Neural Network Architectures, PyTorch, Object Detection Algorithms, and Computer Vision Applications for Python Coders (English Edition) 1st Edition, Kindle Edition, ISBN-13: 978-9390684687 ISBN-10: 9390684684.

[9] Redmon, Joseph & Divvala, Santosh & Girshick, Ross & Farhadi, Ali. (2016). You Only Look Once: Unified, Real-Time Object Detection. 779-788. 10.1109/CVPR.2016.91.

[10] Li, Xueqi & Liu, Qing & Liu, Tongcai & Wang, Jingbo. (2021). Research on YOLO Model and Its Application in Fault Status Recognition of Freight Trains. 10.1007/978-3-030-78615-1_13.

[11] Baca, Arnold. (2021). Computer Vision. 10.4324/9781003157007-21.

[12] Bhalley, Rahul. (2021). Computer Vision 10.1007/978-1-4842-6330-3_6.

[13] Finlay, Janet & Dix, Alan. (2020). Computer vision. 10.1201/9781003072485-8.

[14] Sharma, Lavanya. (2018). Object Detection using Background Subtraction.

[15] Shaikh, Soharab & Saeed, Khalid & Chaki, Nabendu. (2014). Moving Object Detection Using Background Subtraction. 10.1007/978-3-319-07386-6_3.

[16] Brill, Frank & Erukhimov, Victor & Giduthuri, Radhakrishna & Ramm, Stephen. (2020). Background subtraction and object detection. 10.1016/B978-0-12- 816425-9.00013-9.

[17] Obukhov, Anton. (2011). Haar Classifiers for Object Detection with CUDA. 10.1016/B978-0-12-384988- 5.00033-4.

[18] Useche Murillo, Paula & Moreno, Robinson & Baquero, Javier. (2020). Algorithm of detection, classification and gripping of occluded objects by CNN techniques and Haar classifiers. International Journal of Electrical and Computer Engineering (IJECE). 10. 4712. 10.11591/ijece.v10i5.pp4712-4720.

[19] Kanburoğlu, Ali Buğra & Tek, F.. (2018). A Haar Classifier Based Call Number Detection and Counting Method for Library Books. 10.1109/UBMK.2018.8566314.

[20] Gothane, Suwarna. (2021). A Practice for Object Detection Using YOLO Algorithm. International Journal of Scientific Research in Computer Science, Engineering and Information Technology. 268-272. 10.32628/CSEIT217249.

[21] Padma Reddy, Monalika & A, Deeksha. (2021). Mulberry leaf disease detection using YOLO.

[22] Hatab, Muhieddine & Malekmohamadi, Hossein & Amira, Abbes. (2021). Surface Defect Detection Using YOLO Network. 10.1007/978-3-030-55180-3_37.

[23] Hiremath, Rajshekar & Malshikare, Komal & Mahajan, Manish & Kulkarni, Radhika. (2021). A Smart App for Pothole Detection Using Yolo Model. 10.1007/978-981-15-8354-4_16.

[24] Lohit, G. & Sampath, Nalini. (2020). Multiple Object Detection Mechanism Using YOLO. 10.1007/978-981- 15-1097-7_48.

[25] Zhang, Ling & Chen, Siping & Wang, Tianfu & Liu, Zhuo. (2012). Automatic Facial Expression Recognition Based on Hybrid Features. Energy Procedia. 17. 1817- 1823. 10.1016/j.egypro.2012.02.317.

[26] R. Ranjan, S. Sankaranarayanan, C. D. Castillo and R. Chellappa, "An All-In-One Convolutional Neural Network for Face Analysis," 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017), 2017, pp. 17-24, doi: 10.1109/FG.2017.137.

[27] Jalali, Amin & Mallipeddi, Rammohan & Lee, Minho. (2017). Sensitive Deep Convolutional Neural Network for Face Recognition at Large Standoffs with Small Dataset. Expert Systems with Applications. 87. 10.1016/j.eswa.2017.06.025.

[28] Saez Trigueros, Daniel & Meng, Li & Hartnett, Margaret. (2017). Enhancing Convolutional Neural Networks for Face Recognition with Occlusion Maps and Batch Triplet Loss. Image and Vision Computing. 79. 10.1016/j.imavis.2018.09.011.

[29] Cordeiro, Filipe & Chevtchenko, Sergio & Vale, Rafaella & Macario, Valmir. (2018). A Convolutional Neural Network with Feature Fusion for Real-Time Hand Posture Recognition. Applied Soft Computing. 10.1016/j.asoc.2018.09.010.

[30] Guo, G., Wang, H., Yan, Y., Zheng, J., & Li, B.

(2020). A Fast Face Detection Method via Convolutional Neural Network. ArXiv, abs/1803.10103.

[31] C. Ding and D. Tao, "Trunk-Branch Ensemble Convolutional Neural Networks for Video-Based Face Recognition," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 40, no. 4, pp. 1002-1014, 1 April 2018, doi: 10.1109/TPAMI.2017.2700390.

[32] Ahsan, Mominul & Based, Md & Haider, Julfikar & Kowalski, Marcin. (2021). COVID-19 Detection from Chest X-ray Images Using Feature Fusion and Deep Learning. Sensors. 21. 10.3390/s21041480.

[33] S. Iqbal et al., "Prostate Cancer Detection Using Deep Learning and Traditional Techniques," in IEEE Access, vol. 9, pp. 27085-27100, 2021, doi: 10.1109/ACCESS.2021.3057654.

[34] Kang, Jaeyong, Zahid Ullah, and Jeonghwan Gwak. 2021. "MRI-Based Brain Tumor Classification Using Ensemble of DeepFeatures and Machine Learning Classifiers" Sensors 21, no. 6: 2222. https://doi.org/10.3390/s21062222

[35] B. Maurya, S. Hiranwal and M. Kumar, "A Review on Liver Cancer DetectionTechniques," 2020 5th IEEE International Conference on Recent Advances andInnovations in Engineering (ICRAIE), 2020,pp. 1-5, doi: 10.1109/ICRAIE51050.2020.9358362.

[36] Sharma, Niharika & Mishra, Prakhar & Sadhana, K.S. & Nithyakani, P.. (2018). Segmenting and classifying MRI images for brain tumors using CNN. International Journal of Engineering and Technology (UAE). 7. 146-149. 10.14419/ijet.v7i3.29.18545.