Study on Object Detection Using Deep Learning Models

ANUJ KUMAR YADAV[1], DR. NARESH CHAUHAN[2]
[1] *Student, M. Tech (CE), J.C Bose University of Science and Technology, YMCA, Faridabad, India*
[2] *Professor, CSE, J.C Bose University of Science and Technology, YMCA, Faridabad, India*

Abstract— Object Detection is a technique related to detecting objects an object of a certain class (such as a person, building, or car) in digital images or videos. The machine will process the image or video and will produces output for each detected object along with the appropriate label with classes. The engine will also display a bounding box on the detected object and will display the accuracy value of detection. However, this study evaluates the various models used in object detection using Deep Learning Techniques. Therefore, as per various researches in past and adhered the scheme observed that the Convolution Neural Networks is an effective model to identify object's accurately and swiftly.

Indexed Terms-- Object Detection, Computer Vision, Deep Learning, Neural Networks.

I. INTRODUCTION

Surveillance camera is a camera that serves to assist the process of monitoring an area within the scope of the digital image it takes. Surveillance cameras provide surveillance results in the form of images and videos that can be accessed in real time while saving them into storage media. The stored digital images are generally the entire recording, regardless of the activities that occur or the existence of the objects contained in them, including human objects. Under such conditions, the recorded data produced by the surveillance camera will result in a waste of memory on the storage media. These limitations can also be significant depending on the particular situation. For example, in real time conditions, the surveillance camera can only send video to the monitor to be monitored by the security guard, so the function of the surveillance camera will depend on the response of the guard (user). If the guard (user) happens to be careless, it is too late to take preventive action or handling an incident. In addition, the recordings that have been stored in memory are also often used as important references by the authorities in the process of reviewing human activities that have occurred in them. This is what makes the existence of surveillance cameras play an important role in helping to monitor security in an area, but its application is still not optimal in terms of functionality and memory allocation [1].

Therefore, we need a method that can make the surveillance camera active so that it can function efficiently and optimally. Along with the rapid development of information technology, now the application of technology in the field of Computer Vision [2] is able to overcome these problems.

Computer Vision is a part of artificial intelligence [3] that focuses on training computers to interpret and understand the visual world. Machines can identify and classify objects accurately from images and Deep Learning [4, 5] models and then provide feedback on these images. In embedded systems, one way that is often used to overcome these problems is to use cloud computing services. However, these services require additional costs and have the potential to experience data latency.

Another alternative is to use a special hardware accelerator such as the Movidius Neural Compute Stick (NCS) device. Intel Movidius NCS is powered by a power-efficient and high-performance Vision Processing Unit (VPU) so that it can provide good computing power to run Deep Learning models on the system locally [6-8]. Several studies that have been done previously related to the problems that have been described include a study entitled "Object Tracking and Suspicious Activity Identification during Occlusion" by [9] which detects human objects and processes suspicious conditions based on differences in grayscale levels. the average of the processed data. Other related research was also conducted by [10] in 2014, who built a system to Motion Tracking of

Humans under Occlusion Using Blobs method with the results of the GMM (Gaussian Mixture Model).

II. LITERATURE REVIEW

- Object Detection

When humans glance at an image, the human brain can immediately recognize the object in the image, the location of the object, and the conditions of interaction that occur. The human visual system is fast and accurate, enabling us to perform complex tasks. Fast and accurate object detection algorithms will allow computers to do similar things to the potential to complete common tasks. Detection of objects in digital image processing is a process used to determine the presence of certain objects in a digital image.

The detection process can be carried out by various methods which generally read the features of all objects in the input image. The features of the objects in the input image will be compared with the features of the model used or template. The comparison results can be used to determine whether an object is detected as the intended template or not. The object detection system needs to train and test datasets with bounding boxes and labels for the class per each object for the recognition process. To achieve this goal, there are many datasets for generating Deep Learning models such as Pascal-VOC [11].

- Literature References

1. There is a research conducted by [12] R. Chauhan, K. K. Ghanshala and R. C. Joshi, 2018 which applies the Deep Learning method in classifying 2D images with Convolutional Neural Network (CNN) algorithm. In this study using an input shape measuring 32x32, the value of learning rate 0.001, the filter size is 3x3, the number of epochs is 50, whereas the MNIST dataset based 60,000 train images and 10,000 test images along with CIFAR-10 is the dataset used for object detection which is labeled a subset of 80 million tiny images are used. However, the dataset has 60,000 32x32 pixel color images with 10 classes (airplane, automobile, bird, cat, deer, dog, frog, horse, ship, truck). Each class has 6000 images. The train batch has 50,000 images and test batch has 10,000 images. The test batch for each class has 1000 images which are randomly selected. Based on the results of existing research obtained the calculated accuracy on MNIST is 99.6% and on CIFAR-10 is 80.17%.

2. There is a study [13] that detects real time human objects using the Histogram of Oriented Gradients (HOG) method conducted by Hou, Beiping & Wen, Zhu. (2011). The resulting HOG feature will be processed using the Support Vector Machine (SVM) method to determine whether the detected feature is a human feature or not. The data that is processed in this study is a video recording based on HOG features are calculated as classified eigenvectors. Thereafter test results are categorized into predefined groups of humans and other objects by using SVM classifier , the system can detect the people on visitor videos using the proposed method with an effectiveness and accurately.

3. The research [14] carried out by Pranamurti, H & Murti, A & Setianingsih, C. (2019) is to design a fire detector using image processing and processed by Raspberry Pi. The images sampled in this study are fire flames i.e match fires and paper burnt flames. The method used in this study is the Haar Cascade Classifier method. The results show the average value of the output displayed by the system, namely in the fire experiment with the distance of 50cm to 100cm the highest accuracy of 100% with the light of 0 lx.

4. Another study [15] entitled "Object Motion Detection and Tracking for Video Surveillance" conducted by Sahasri, M., & Gireesh, C in 2017 used background subtraction and foreground mask sampling methods with a cascade classifier for object detection processes. This research was conducted because of the need by "smart" video surveillance systems for fast, reliable and robust algorithms for detecting moving objects. This study also takes samples by calculating intersections on a number of frames that have been background-subtracted in a certain time period.

5. By [16] Kumar, A., Zhang, Z.J. & Lyu, H investigates and build up an object detector algorithm with deep learning neural networks for

detecting the objects from the images. The investigations apply am enhanced SSD algorithm alongside through multilayer convolutional network to accomplish high accurateness in factual instance for the recognition of the objects. The presentation of proposed algorithm by researchers is excellent in motionless images and moving videos. The accurateness of the projected representation is extra than 79.8%. The training time for this model is about 5–6h. These convolutional neural networks extract feature information from the image and then perform feature mapping to classify the class label. The prime objective of our algorithm is to use the best aspect ratios values for selecting the default boxes so that we can improve SSD algorithm for detecting objects.

6. Teja [17] in 2018 conducted a study to detect and track the movement of some objects during occlusion and activate warning if there is an object that is visible for a long time in the scope camera. The existing system uses the concept of video framing into several types 2-dimensional image and then apply digital image processing techniques to the image to detect the motion of some objects on the video. Multiple object detection obtained based on differences in the average grayscale level of the processed data.

7. In the scientific article by Rajaraman Sivaramakrishnan [18] (2015), with the title "Pre-trained Convolutional Neural Networks as feature extractors to improve the detection of the malaria parasite in thin blood images", the main objective was to find a model whose evaluation show capacity for assertiveness of deep learning based on Convolutional neural networks, for the selection and classification of microscopic images. The development methodology that he chose to carry out was the cell segmentation algorithms, allowing to obtain values of sensitivity and specificity based on the procedure carried out. It was concluded that the personalized model was able to meet the expected expectations, achieving 81% sensitivity and 83% specificity compared to the diagnostic accuracy that was evaluated.

8. In [19] the scientific article by Wu, Chunyan and Wang, Xuefeng (2013), with the title "Preliminary investigation on the identification system of anthracnose and powdery mildew of sandalwood leaves based on image processing", it was sought to identify the most accurate system for the preliminary diagnosis of sandalwood diseases through fragmentation, analysis of characteristics such as color and texture of digital images and their respective categorization. The development methodology used is based on convolutional Neural Networks (input layer, hidden laye layer, output layer) through learning algorithms (propagation and segmentation). It was concluded that using neural networks they can respond optimally in image recognition models to identify diseases.

9. In [20] the scientific article by Abdullah-Al Nahid, Menrabi Mohamad and Yinang Kong (2018), with the title of "Histopathological Classification of Breast Cancer Imaging by Deep Neural Network Techniques Guided by Local Grouping" with the main objective is to contribute in the Histopathological classification of breast cancer and thus be able to help in decision making. Likewise, the development method he carried out was a convolutional neural network performing biomedical image analysis, classifying it using techniques guided by statistical information derived from images. In the same way, it was concluded that the system provides a great facility for decision-making since if it were not present, a whole process would be carried out depending on whether the state is malignant or benign to delegate a specialist.

10. In [21] the scientific article, Raith Stefan, Per Eric, Gut Cristine, Edelhoff Daniel and Fischer Horst (2017), with the title "Artificial neural networks as a powerful numerical tool to classify the specific characteristics of a tooth based on 3D scan data " with the aim of being able to give precision to the classification of tooth images with respect through neural networks so that it can be a contribution in the design of treatment and restorations in the area of Dentistry. In the same way, the neural network development method uses Python as the programming language with a package called

Bryan and using the so-called backpropagation algorithm

11. In [22] the scientific article by Burcu Yavuz, Tuba Yildiz and Ziynet Pamuk (2014), with the title "Comparison and Regression tree classifiers used with the Backpropagation algorithm to diagnose Anemia" with the aim of developing a method to improve classification performance by hematological parameters. Likewise, the method used was the selection algorithm in relation to the Gini algorithm based on the regression tree together with the backpropagation algorithm for the diagnosis of Anemia. In the same way, it was possible to conclude that the aforementioned method provides help to doctors and can even be used for various disease diagnostic studies. From the article mentioned, the application of the Backptopagation algorithm is taken as experience, since to improve the classification performance, go through an error propagation stage.

12. In [23] the scientific article by Hsu-Hao Yang, Mei-Ling, Huang, Shin-Wei Yang (2015), with the title "Integration of Auto-Associative Neural Networks with Hotelling T2 Control Charts for Wind Turbine Fault Detection" Its main objective is to be able to contribute to the constant failures of a wind turbine through CAD and neural networks. Likewise, the methodology used is K-means (grouping method so that value data can be processed and grouping results can be verified, associative neural networks are also applied that transform the data to be distributed. In the same way, was able to conclude that the aspect of precision that was required can be improved by selecting attributes and even indicates that operational ranges can be added to improve the detection of values.

13. In [24] the scientific article by Jusman Yessi, Siew Cheok and Noor Azuan (2014), with the title "Intelligent detection systems for cervical cancer", with the main objective of being able to analyze intelligent systems that focus on diagnosing cervical cancer. cervix by comparing images. Likewise, the methodology is based on the review of two scientific databases for general visions of intelligent systems among others. In the same way, it was concluded that there are various classification tools (neural networks, logistic regression, decision tree, among others) that generate better results for classifying cervical precancerous data, since a real-time system can even be built.

14. In [25] the scientific article by Jitendra Virmani, Vinod Kalra Naveen and Niranjan Khandelwal (2014), with the title of "Neural Network Ensemble-Based CAD System for B-mode Ultrasound Focal Liver Lesions" which has as its main objective to provide a Diagnosis of focal liver lesions using a system based on a set of neural networks. Also, the method it uses is a neural network ensemble using a CAD diagnostic system consisting of extraction functions and a step classification module.

15. In [26] the scientific article by Shi Yun, Xianfeng Wang, Shanwen Zhang and Zhan Chuanlei (2015), with the title of "PNN-based crop disease recognition with leaf image characteristics and meteorological data" whose main objective is to propose a method an automatic recognition of crop diseases. Likewise, the development method that was applied was the statistical characteristics of texture, color and shape through the images obtained, a probabilistic neural network classifier was also chosen so that the accuracy of the classification can be evaluated. In the same way the results of a set of images of diseased leaves can be recognized by the method mentioned even the recognition rate accuracy was greater than 90%.

III. CONCLUSION

In this study, various methods are referred and adhered in literature for solving the problem of detection and classification of objects in images, which belongs to the branch of Deep Learning and which are called Neural Networks. Various techniques and results have been given to specific problems of detection in cases with a single object to classify and in others with images in which it is interesting distinguish several different objects and their different locations. This has required a detailed study of the architecture of Convolutional Neural Networks, Histogram of Oriented Gradients and Support Vector Machine of the

parameters that affect their training, thus modifying the results that is possible obtain in a classification of the type described. As per the literature review to obtain the most optimal solution possible from the data and tools that were available the CNN is the recommended and effective technique to detect objects accurately and swiftly.

REFERENCES

[1] **E**oghan Casey. 2011. Digital Evidence and Computer Crime: Forensic Science, Computers, and the Internet (3rd. ed.). Academic Press, Inc., USA.

[2] Rasche, Christoph. (2019). Computer Vision.

[3] Atallah, Sam. (2021). Artificial Intelligence and Computer Vision. 10.1007/978-3-030-49100-0_31.

[4] Okut, Hayrettin. (2021). Deep Learning.

[5] Yagawa, Genki & Oishi, Atsuya. (2021). Deep Learning. 10.1007/978-3-030-66111-3_3.

[6] Hamid, Muhammad & Zaman, Fadhlan. (2019). Hand Gesture Recognition Using Movidius Neural Compute Stick. 510-514. 10.1109/ICSEngT.2019.8906357.

[7] Kaarmukilan, s.P. & Hazarika, Anakhi & Thomas K, Amal & Poddar, Soumyajit & Rahaman, Hafizur. (2020). An Accelerated Prototype with Movidius Neural Compute Stick for Real-Time Object Detection. 1-5. 10.1109/ISDCS49393.2020.9262996.

[8] Xing, Yannan & Kirkland, Paul & Caterina, G. & Soraghan, John & Matich, G.. (2018). Real-Time Embedded Intelligence System: Emotion Recognition on Raspberry Pi with Intel NCS: 27th International Conference on Artificial Neural Networks, Rhodes, Greece, October 4-7, 2018, Proceedings, Part I. 10.1007/978-3-030-01418-6_78.

[9] Teja, Ravi & Nayar, Raunaq & Sreedevi, Indu. (2018). Object Tracking and Suspicious Activity Identification during Occlusion. International Journal of Computer Applications. 179. 29-34. 10.5120/ijca2018916117.

[10] Mariappan, Sivarathinabala & Murugappan, Abirami. (2014). Motion Tracking of Humans under Occlusion Using Blobs. 27. 251-258. 10.1007/978-3-319-07353-8_30.

[11] Vicente, Sara & Carreira, J. & Agapito, Lourdes & Batista, Jorge. (2014). Reconstructing PASCAL VOC. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 41-48. 10.1109/CVPR.2014.13.

[12] R. Chauhan, K. K. Ghanshala and R. C. Joshi, "Convolutional Neural Network (CNN) for Image Detection and Recognition," 2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC), 2018, pp. 278-282, doi: 10.1109/ICSCCC.2018.8703316.

[13] Hou, Beiping & Wen, Zhu. (2011). Fast Human Detection Using Motion Detection and Histogram of Oriented Gradients. JCP. 6. 1597-1604. 10.4304/jcp.6.8.1597-1604.

[14] Pranamurti, H & Murti, A & Setianingsih, C. (2019). Fire Detection Use CCTV with Image Processing Based Raspberry Pi. Journal of Physics: Conference Series. 1201. 012015. 10.1088/1742-6596/1201/1/012015.

[15] Sahasri, M., & Gireesh, C. (2017). Object Motion Detection and Tracking for Video Surveillance. International Journal of Engineering Trends and Technology, (April), 161–164.

[16] Kumar, A., Zhang, Z.J. & Lyu, H. Object detection in real time based on improved single shot multi-box detector algorithm. J Wireless Com Network 2020, 204 (2020). https://doi.org/10.1186/s13638-020-01826-x

[17] Teja, Ravi & Nayar, Raunaq & Sreedevi, Indu. (2018). Object Tracking and Suspicious Activity Identification during Occlusion. International Journal of Computer Applications. 179. 29-34. 10.5120/ijca2018916117.

[18] Rajaraman, Sivaramakrishnan & Antani, Sameer & Poostchi, Mahdieh & Silamut, Kamolrat & Hossain, Md & Maude, Richard & Jaeger, Stefan & Thoma, George. (2018). Pre-trained convolutional neural networks as feature extractors toward improved malaria parasite detection in thin blood smear images. PeerJ. 6. 10.7717/peerj.4568.

[19] Wu, Chunyan & Wang, Xuefeng. (2017). Preliminary research on the identification system for anthracnose and powdery mildew of sandalwood leaf based on image processing. PLoS ONE. 12. 10.1371/journal.pone.0181537.

[20] Nahid, Abdullah & Mehrabi, Ali & Kong, Yinan. (2018). Histopathological Breast Cancer Image Classification by Deep Neural Network Techniques Guided by Local Clustering. BioMed Research International. 2018. 1-20. 10.1155/2018/2362108.

[21] Raith, Stefan & Vogel, Eric & Anees, Naeema & Keul, Christine & Güth, Jan-Frederik & Edelhoff, Daniel & Fischer, Horst. (2016). Artificial Neural Networks as a powerful numerical tool to classify specific features of a tooth based on 3D scan data. Computers in Biology and Medicine. 80. 10.1016/j.compbiomed.2016.11.013.

[22] Çarkli Yavuz, Burcu & K.Y., Tuba & Yurtay, Nilüfer & Pamuk, Ziynet. (2014). Comparison Of K Nearest Neighbours And Regression Tree Classifiers Used With Clonal Selection Algorithm To Diagnose Haematological Diseases. AJIT-e: Online Academic Journal of Information Technology. 5. 7-20. 10.5824/1309-1581.2014.3.001.x.

[23] Yang, Hsu-Hao & Huang, M L & Yang, Shih-Wei. (2015). Integrating Auto-Associative Neural Networks with Hotelling T2 Control Charts for Wind Turbine Fault Detection. Energies. 8. 12100-12115. 10.3390/en81012100.

[24] Jusman, Y., Ng, S. C., & Abu Osman, N. A. (2014). Intelligent screening systems for cervical cancer. TheScientificWorldJournal, 2014, 810368. https://doi.org/10.1155/2014/810368

[25] Virmani, Jitendra & Kumar, Vinod & Kalra, Naveen & Khandelwal, Niranjan. (2014). Neural Network Ensemble Based CAD System for Focal Liver Lesions from B-Mode Ultrasound. Journal of digital imaging. 27. 10.1007/s10278-014-9685-0.

[26] Shi, Y. & Wang, X.F. & Zhang, S.W. & Zhang, C.L.. (2015). PNN based crop disease recognition with leaf image features and meteorological data. International Journal of Agricultural and Biological Engineering. 8. 60-68. 10.3965/j.ijabe.20150804.1719.