Survey on Object Recognition Techniques using Machine Learning

Pavan S Patil[1], Santhoshkumar V[2], Udaykumar V[3], Vishwas K K[4]
[1234] Electronics and Communication Department, KSIT College
Dr. B Sudarshan[5], Associate Professor
KSIT, Bengaluru, India

***Abstract*---Object Recognition is one of the most exciting areas in machine learning right now. Recognize objects like faces or cats are not so difficult, but recognizing arbitrary objects within a larger image set has been the difficult by artificial intelligence. The real surprise is that human brains recognize objects so well and effortlessly convert photons bouncing off objects into a spectacularly rich set of information about the objects around us. Machine learning still struggles with these simple tasks, but in the past few years, it's gotten much better.**

***Keywords*—Deep Neural Network (DNN), Convolutional Neural Network (CNN), Object recognition, Machine Learning, Deep Learning**

I.INTRODUCTION

Object Recognition technology has seen a quite good adoption rate in various and diverse industries. It helps in safe navigation of self-driving vehicles through traffic, spots violent behavior in a crowded place, it can used by sports teams for analyzing and build scouting reports, quality control of parts in manufacturing can be monitored and many other things. Deep learning is adopted so that brain neural network functions to perceive the data, such as image, sound can be simulated artificially. Deep learning by ImageNet has made an huge amount of progress toward object recognition by, collecting and processing those data sets. This paper involves analyzing various machine learning techniques for object recognition, so that different methodologies and algorithms used for object recognition can be achieved.

II.LITERATURE SURVEY

The paper [1] proposes deep learning approach in the recognition of objects in the historical building photographs of the town Trnava. It uses Deep learning architectures based on convolutional neural networks (CNN) for object recognition tasks. Cascade of convolution layers and activation functions are used to improve architecture. It is very important to setup of the number of layers and the number of neurons in each layer. TRNAVA LeNet 10 model was built and trained the purpose. This model is based on the dataset of 460 training images and 140 validation images which is the ratio of 3:1, images are of dimensions 28x28 pixels and image type used was color, image encoding was jpg. The model successfully recognized the right object in the photograph of historical building in Trnava. The proposed model gained 98.88% prediction accuracy.

The paper [2] proposes deep learning methods for facial expression recognition instead of hand-crafted features. Two kinds of deep networks such as deep neural network (DNN) and convolutional neural network (CNN) are used to solve recognition problems. The deep networks were developed using CUDA supported deep learning toolkits such as Caffe and CudaConvnet2 for high speed. Also, for implementing Haar-like face detection algorithm, they used OpenCV library. The images were cropped and resized to 64×64. Then, the 327 face images were divided to 10 groups, and then used one group for training and nine groups for test. The recognition results were good for 6 emotions, but the recognition rate of disgust label was poor. Because, the number of training images of disgust label in FER 2013 database was only 547. The DNN has the possibility of overfitting.

The paper [3] proposes considerable improvement in object detection and tagging using convolutional neural networks has given way to accurate yet complex methods, which can identify objects in real-time. However, the growth in the area of implementing the algorithms on low powered portable devices has been relatively slow, aims to converge the fields of computer vision and robotics, focusing on implementation of image description applications on an embedded system platform. The objects in the image are restricted to a fixed number, specific to data set used for training the model. According to Shaoqing Ren et al, the introduction of Region Proposal Network (RPN) allows sharing of whole image convolutional features with the network, thus, providing near cost-free region proposals. Wherein, region proposal technique is used to guide the algorithm in order to locate objects residing in an image. Secondly, execution of this method in our system allows the system to be computationally efficient and customized to run on low-powered machines.

The paper [4] proposes an object localization method to boost the performance of current object recognition techniques, utilizes the image edge information as a clue to determine the location of the objects. The Generic Edge Tokens (GETs) of the image are extracted based on the perceptual organization elements of human vision. These edge tokens are parsed according to the Best First

Search algorithm to fine-tune the location of objects, where the objective function is the detection score returned by the Deep Convolutional Neural Network. Applying the BFS to the object localization and its search space, the search space is a set of edge elements whose overlaps with the current candidate object is greater than zero. Testing the model in real time proved to be more efficient than the RCNN, also with scope for further development by improving the object localization by using a combination of the image edge, color and texture information, and the learned features of the image.

The paper [5] proposes a method to develop an interactive application in order to detect objects from videos, upon user input, it is also able to detect the particular object being shown at that instant on the screen. A sequential frame extraction method of videos and also deep learning approach of Convolutional Neural Networks along with Fully Connected Neural Networks is used for this task which gives an accuracy rate of 77%. Even when the object is somewhat distorted, translated, rotated or partially obstructed from view, it can be easily detected by humans, the task of computer vision is still quite challenging. Videos are made of frames synchronized with some playback audio, taking advantage of the fact that videos consist of frames, the analysis of the video can be made in much more detail by examining the object present in the frame images themselves, running the classifier and thereby get probabilities for different classes and hence classify the genre and also detect any object in the video. The operational accuracy of this model is improved by increasing the number of datasets and improving the hardware configuration so that the object classification can be done over a wider range of classes and in a faster way.

The paper [6] proposes an approach to use the concept of deep learning with the convolutional neural networks in identifying the objects. It uses the input video to give the output with the set of identified objects. The convolutional neural network works give the confidence score for each of the objects. It uses Single Shot Multibox Detector which identifies the multiple objects at a time with the help of convolutional network and also it had high accuracy rate. It uses Hard Negative Mining and Non-Maximum Suppression, these are used to increase the confidence score of the object and to generate only one detection for each objects respectively. Hence these helps picking highest score and avoids in multiple detection of object. Hence neural networks along with deep residual networks increases the computational speed and accuracy in identification of objects.

This paper [7] proposes an object-based coding method for very low bit-rate channels, using a method based on motion estimation with a block-based moment-preserving edge detector. The global motion components are the most widely used object based coding methods. The problem with global motion components are prediction error is large, even after motion compensation using the discrete cosine transform (DCT), this happens when images contain rapid moving objects and noise. Furthermore, it cannot result in small prediction error if the segmented objects consist of sub-objects that move through different directions. The technique proposed in this paper involves a hybrid object-based video-coding approach that retains the relative advantages of both the object-based and block-based coders while minimizing the drawbacks of both which is segmenting moving objects from video sequences and representing objects compactly by visual-pattern approximations of the boundary. It is achieved by detecting the line edge from a square block using the moment preserving edge detector. Due to high computational complexity is required for motion estimation using block matching. To reduce the complexity of this a fast block-matching method is used which is based on the Visual patterns. Results show that the proposed method gives good efficiency in terms of the subjective quality, the peak signal-to-noise ratio (PSNR), and the compression ratio.

This paper [8] proposes a new object detection algorithm which uses a convolutional network with a convolution kernel of the Network in Network (NiN) type which allows the use of massively parallel processing. Non-linear approach of the convolution kernel allows to provide a large stride and to abandon the pooling only when it is used in the form of a fully connected network. The simultaneous localization of objects on an image and their recognition is knows as Detection. Detector can operate with images of arbitrary sizes. The algorithm has a high computational efficiency, so when processing HD frame on a single CPU core, the operating time may increase up to 300 ms. Massively parallel data processing on GPU is result of high degree of uniformity of network operations, which is likely to reduce the operating time to less than 10 ms. The proposed algorithm is robust to small overlaps and the average quality of images of detected objects. It is an end-to-end learner model and output is delimited by boundaries and classes of objects throughout an image. An open access image database is used to evaluate the algorithm. This algorithm is not limited to the use of one type of objects, it can simultaneously detect a mixture of objects. The algorithm proposed achieves high speed image processing, and the efficiency is comparably higher than the other algorithms.

This paper [9] addresses the problem of online tracking and classification of multiple objects in an image sequence. Solution is to first track all objects in the image without relying on object-oriented prior knowledge, which can be of hand-crafted features or track initialization which is user based. A fast-learning image classifier is used to classify the tracked objects, which is based on a shallow convolutional neural network architecture and also when it is combined with object state information from the tracking algorithm the efficiency object recognition improves. A robust, general purpose object recognition system with the ability to detect and track a variety of object types can be achieved by transferring the prior knowledge from the detection and tracking stages to the classification stage. The system adaptively learns the shape and motion of tracked objects, and apply it to Neovision2 Tower data set, which contains multiple objects. An. evaluation demonstrates that the approach is competitive which make use of object-specific prior knowledge in detection and tracking and it also provides additional practical advantages by virtue of its generality.

This paper [10] proposes Object confirmation essentially is object detection based on an image. Traditional object detection algorithms include three steps: region selection, feature extraction and classification. YOLOv2 reframes object detection as a single regression problem, straight from image pixels to bounding box coordinates and class probabilities simultaneously by running a single deep convolutional neural network on the image. the low-layer filters extract the detail texture information of objects while the high-layer filters extract the semantic information, multi-feature fusion has become a new trend in the deep convolutional neural network design in recent years. The intelligent radar perimeter security system incorporates the high sensitivity of radar detection and the high accuracy of object confirmation.

III. GAPS ANALYSIS

In paper [1] the TRNAVA LeNet 10 model gained the average prediction accuracy of more than 81.41% and 18.59% was the rate of unsuccessful prediction, as the hardware used has limitations, especially graphics memory.

In paper [2] the Deep neural network and convolutional neural network models would have trained with more test data, such that it would have been while evaluating to achieve good accuracy of more than 90%.

In paper [3] the objects to be recognized is heavily dependent upon the data set used for training the model and its extracted weights, it limits the developed system to detect only particular objects which were included in the data set. This limitation can be overcome by including more different data sets and feeding the system with updated weights.

In paper [4] there is a possibility for improving the object localization by using a combination of the image edge, colour and texture information, and the learned features of the image by using some of the deep learning techniques.

In paper [5] the model was trained with only 20 object classes. Time taken for training was too much, the number of classes couldn't be increased more because of hardware restrictions. Each of the object class has 400 training images and 250 validation images. Thus, the model has been trained over 8000 training images and tested over 5000 testing images.

In paper [6] Hard negative mining results in very slow process for creating lots of negative or false samples. Non-maximum suppression performance is affected by objects that are highly overlapped with each other, and its localization depends solely on the highest scored detection rate for accuracy.

In paper [7] using the proposed method has the significant differences that images cannot always be explained by the gain in PSNR, the reconstructed images include much visible distortion when the PSNR is low. To keep the PSNR values of the reconstructed images high it is difficult to maintain a video sequence at very low bit rates.

In paper [8] it has the drawbacks of the presence of false detections in the algorithm which leads to including one class classification algorithms which leads to abandon the use of real images for training and shift to using modeled objects.

In paper [9] it has the limitations that the system does not know the extent of a single object. It simply associates a set of observations like in shape, position and velocity with a single object. This results in both one false positive and many false negatives in certain object recognition.

In paper [10] under unstable weather conditions, the intelligent radar perimeter security system cannot achieve effective warning in complex scenarios. False alarm rate can be further reduced by introducing the object confirmation module. During extreme weather conditions, false alarms detected by the radar would be increased.

IV.CONCLUSION

Object recognition is one upcoming and exciting areas in machine learning. A well- known application of object detection is face detection, that is used in almost all the mobile cameras. But the major setback in all these papers are real time application and accuracy rate achieved, these can be overcome by Inception Model advanced architecture of ImageNet. These systems can be integrated with other tasks such as pose estimation where the first stage in the pipeline is to detect the object, and then the second stage will be to estimate pose in the detected region. It can be used for tracking objects and thus can be used in robotics and medical applications. Thus, this problem serves a multitude of applications. Deep learning models are the best considered for object detection because of training time, low latency, faster evaluation, etc.

V. REFERENCES

[1] Bezak, P. (2016, September). *Building recognition system based on deep learning*. In 2016 Third International Conference on Artificial Intelligence and Pattern Recognition (AIPR) (pp. 1-5). IEEE.

[2] Jung, H., Lee, S., Park, S., Kim, B., Kim, J., Lee, I., & Ahn, C. (2015, January). *Development of deep learning-based facial expression recognition system*. In 2015 21st Korea-Japan Joint Workshop on Frontiers of Computer Vision (FCV) (pp. 1-4). IEEE.

[3] Tenguria, R., Parkhedkar, S., Modak, N., Madan, R., & Tondwalkar, A. (2017, April). *Design framework for general purpose object recognition on a robotic platform*. In 2017 International Conference on Communication and Signal Processing (ICCSP) (pp. 2157-2160). IEEE.

[4] Etemad, E., & Gao, Q. (2017, September). *Object localization by optimizing convolutional neural network detection score using generic edge features*. In 2017 IEEE International Conference on Image Processing (ICIP) (pp. 675-679). IEEE.

[5] Mazumdar, M., Sarasvathi, V., & Kumar, A. (2017, August). *Object recognition in videos by sequential frame extraction using convolutional neural networks and fully connected neural networks*. In 2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS) (pp. 1485-1488). IEEE.

[6] Sujana, S. R., Abisheck, S. S., Ahmed, A. T., & Chandran, K. S. (2017, April). *Real time object identification using deep convolutional neural networks*. In 2017 International Conference on Communication and Signal Processing (ICCSP) (pp. 1801-1805). IEEE.

[7] Cheng, S. C. (2005). *Visual pattern matching in motion estimation for object-based very low bit-rate coding using moment-preserving edge detection*. IEEE transactions on multimedia, 7(2), 189-200.

[8] Alexeev, A., Matveev, Y., & Kukharev, G. (2018, October). *Using a Fully Connected Convolutional Network to Detect Objects in Images*. In 2018 Fifth International Conference on Social Networks Analysis, Management and Security (SNAMS) (pp. 141-146). IEEE.

[9] Wong, S. C., Stamatescu, V., Gatt, A., Kearney, D., Lee, I., & McDonnell, M. D. (2017). *Track everything: Limiting prior knowledge in online multi-object recognition*. IEEE Transactions on Image Processing, 26(10), 4669-4683.

[10] Yang, L., Wang, L., & Wu, S. (2018, April). *Real-time object recognition algorithm based on deep convolutional neural network*. In 2018 IEEE 3rd International Conference on Cloud Computing and Big Data Analysis (ICCCBDA) (pp. 331-335). IEEE.