ISSN: 2350-0328
International Journal of AdvancedResearch in Science, Engineering and Technology
Vol. 10, Issue 9, September 2023

Object Detection System

Prof. Aziz Makandar, Marnur Rashmi Sharanappa

Professor, Department of Computer Science, Karnataka State Akkamahadevi women's University, Vijayapura Karnataka, India
P.G. Student, Department of Computer Science, Karnataka State Akkamahadevi women's University, Vijayapura Karnataka, India

ABSTRACT: Efficient and accurate object detection has been an important topic in the advancement of computer vision systems. With the advent of machine learning and deep learning techniques, the accuracy for object detection has increased drastically. The project aims to incorporate state-of-the-art technique for object detection with the goal of achieving high accuracy with a real-time performance. In this project, we use a completely machine learning with OpenCV and deep learning-based approach to solve the problem of object detection in an end-to-end fashion. The network is trained on the most challenging publicly available dataset, on which an object detection challenge is conducted annually. The resulting system is fast and accurate, thus aiding those applications which require object detection.

I.INTRODUCTION

A few years ago, the creation of the software and hardware image processing systems was mainly limited to the development of the user interface, which most of the programmers of each firm were engaged in. The situation has been significantly changed with the advent of the Windows operating system when the majority of the developers switched to solving the problems of image processing itself. However, this has not yet led to the cardinal progress in solving typical tasks of recognizing faces, car numbers, road signs, analysing remote and medical images, etc. Each of these "eternal" problems is solved by trial and error by the efforts of numerous groups of the engineers and scientists. As modern technical solutions are turn out to be excessively expensive, the task of automating the creation of the software tools for solving intellectual problems is formulated and intensively solved abroad. In the field of image processing, the required tool kit should be supporting the analysis and recognition of images of previously unknown content and ensure the effective development of applications by ordinary programmers. Just as the Windows toolkit supports the creation of interfaces for solving various applied problems.

Object recognition is to describe a collection of related computer vision tasks that involve activities like identifying objects in digital photographs. Image classification involves activities such as predicting the class of one object in an image. Object localization is refers to identifying the location of one or more objects in an image and drawing an abounding box around their extent. Object detection does the work of combines these two tasks and localizes and classifies one or more objects in an image. When a user or practitioner refers to the term "object recognition ", they often mean "object detection ". It may be challenging for beginners to distinguish between different related computer vision tasks.

II. SIGNIFICANCE OF THE SYSTEM

Object detection is a key technology behind advanced driver assistance systems that enable cars to detect driving lanes or perform pedestrian detection to improve road safety. object detection is also useful in applications such as video surveillance or image retrieval systems.

III. LITERATURE SURVEY

A. Research

In various fields, there is a necessity to detect the target object and also track them effectively while handling occlusions and other included complexities. Many researchers (Almeida andGuting2004, Hsiao-Ping Tsai 2011, Nicolas Papadakis and Aure lie Bugeau 2010) attempted for various approaches in object tracking. The nature of the techniques largely depends on the application domain. Some of the research works which made the evolution to proposed work in the field of object tracking are depicted as follows.

ISSN: 2350-0328
International Journal of AdvancedResearch in Science, Engineering and Technology
Vol. 10, Issue 9, September 2023

B. Object Detection

Object detection is an important task, yet challenging vision task. It is a critical part of many applications such as image search, image auto-annotation and scene understanding, object tracking. Moving object tracking of video image sequences was one of the most important subjects in computer vision. It had already been applied in many computers vision fields, such as smart video surveillance (Arun Hampapur 2005), artificial intelligence, military guidance, safety detection and robot navigation, medical and biological application. In recent years, a number of successful single-object tracking system appeared, but in the presence of several objects, object detection becomes difficult and when objects are fully or partially occluded, they are obtruded from the human vision which further increases the problem of detection. Decreasing illumination and acquisition angle. The proposed MLP based object tracking system is made robust by an optimum selection of unique features and also by implementing the A daboost strong classification method.

C. Background Subtraction

The background subtraction method by Horprasert et al (1999), was able to cope with local illumination changes, such as shadows and highlights, even globe illumination changes. In this method, the background model was statistically modelled on each pixel. Computational colour mode, include the brightness distortion and the chromaticity distortion which was used to distinguish shading background from the ordinary background or moving foreground objects. The background and foreground subtraction method used the following approach. A pixel was

modelledbya4-tuple [Ei, si, ai, bi], where Ei- a vector with expected colour value, si - a vector with the standard deviation of colour value, ai - the variation of the brightness distortion and bi was the variation of the chromaticity distortion of the ith pixel. In the next step, the difference between the background dimageand the current image was evaluated. Each pixel was finally classified into four categories: original 8background, shaded background or shadow, highlighted background and moving foreground object. Liyuan Li et al (2003), contributed a method for detecting foreground objects in non-stationary complex environments containing moving background objects. A Bayes decision rule was used for classification of background and foreground changes based on inter-frame colour co-occurrence statistics. An approach to store and fast retrieve colour cooccurrence statistics was also established. In this method, foreground objects were detected in two steps. First, both the for ground and the background changes are extracted using background subtraction and temporal differencing. The frequent background changes were then recognized using the Bayes decision rule based on the learned colour co-occurrence statistics. Both short-term and long-term strategies to learn the frequent background changes were used. An algorithm focused on obtaining the stationary foreground regions as said by Álvaro Bayona et al (2010), which was useful for applications like the detection of abandoned/stolen objects and parked vehicles. This algorithm mainly used two steps. Firstly, a sub-sampling scheme based on background subtraction techniques was implemented to obtain stationary foreground regions. This detects foreground changes at different time instants in the same pixel locations. This was done by using a Gaussian distribution function. Secondly, some modifications were introduced on this base algorithm such as thresh holding the previously computed subtraction. The main purpose of this algorithm was reducing the amount of stationary foreground detected.

D. Template Matching

Template Matching is the technique of finding small parts of an image which match a template image. It slides the template from the top left to the bottom right of the image and compares for the best match with the template. The template dimension should be equal to the reference smaller than the reference image. It recognizes the segment with the highest correlation as the target. image S and an image T, where the dimension of S was both larger than T, output whether S contains a subset image I where I and T are suitably similar in pattern and if such I exists, output the location of I in S as in Hager and Bellhumear (1998). Schweitzer et al (2011), derived an algorithm which used both upper and lowers bound to detect ‘k’ best matches. Euclidean distance and Walsh transform kernels are used to calculate match measure. The positive things included the usage of priority queue improved quality of decision as to which bound-improved and when good matches exist inherent cost was dominant and it improved performance. But there were constraints like the absence of good matches that lead to queue cost and the arithmetic operation cost was higher. The proposed methods dint use queue thereby avoiding the queue cost rather used template matching. Visual tracking methods can be roughly categorized in two ways namely, the feature-based and region-based method as proposed by Ken Ito and Shigeyuki Sakane (2001). The feature-based approach estimates the 3D pose of a target object to fit the image features the edges, givena3D geometrical model of an object. This method requires much computational cost. Region-based can be classified into two categories namely, parametric method and view-based method. The parametric method assumes a parametric model of the images in the target image and calculates optimal fitting of the model to pixel data in a region.

ISSN: 2350-0328
International Journal of AdvancedResearch in Science, Engineering and Technology
Vol. 10, Issue 9, September 2023

The view-based method was used to find the best match of a region in a search area given the reference template. This has the advantage that it does not require much computational complexity as in the feature-based approach.

IV. METHODOLOGY

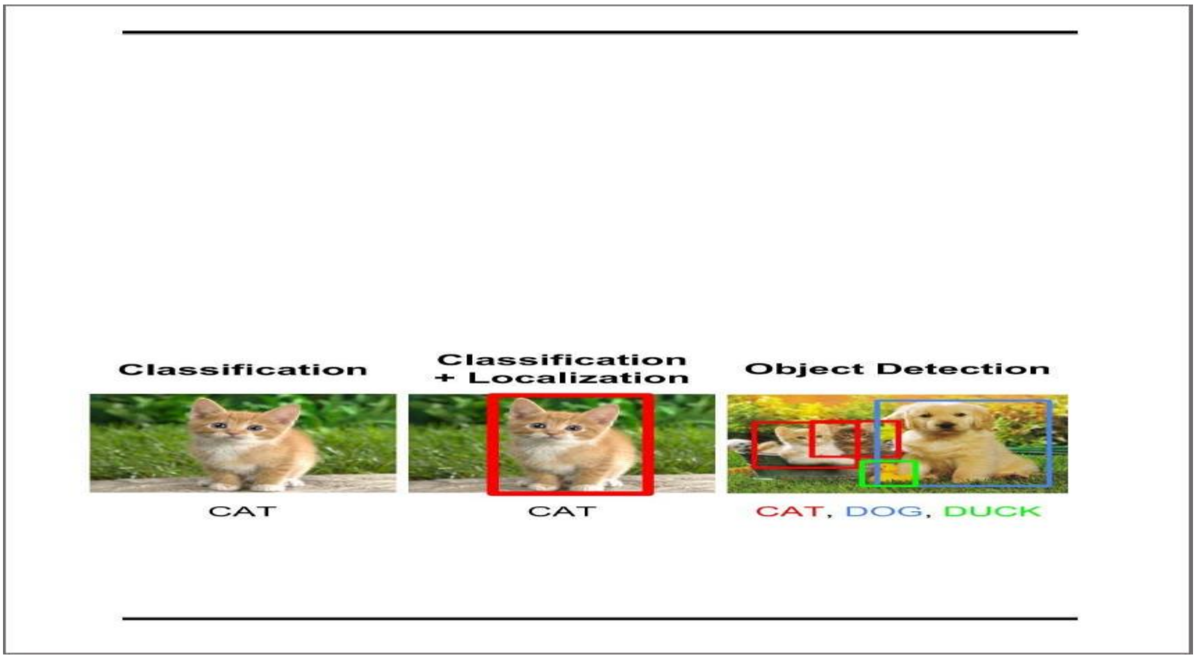
Neural Network Deep learning used by the network has been constantly improving, in addition to the changes in the network structure, the more is to do some tune based on the original network or apply some trick to make the network performance to enhance. The more well-known algorithms of object detection are a series of algorithms based on R-CNN, mainly in the following.

R-CNN

Paper which the R-CNN (Regions with Convolutional Neural Network) is in has been the state-of-art papers in field of object detection in 2014 years. The idea of this paper has changed the general idea of object detection. Later, algorithms in many literatures on deep learning of object detection basically inherited this idea which is the core algorithm for object detection with deep learning. One of the most noteworthy points of this paper is that the CNN is applied to the candidate box to extract the feature vector, and the second is to propose a way to effectively train large CNNs.

SSD

The network used in this project is based on Single shot detection (SSD).
The SSD normally starts with a VGG model, which is converted to a fully convolutional network. Then we attach some extra convolutional layers that help to handle bigger objects.



The output at the VGG network is a 38x38 feature map (conv4 3). The added layers produce19x19, 10x10, 5x5, 3x3, 1x1 feature maps. All these feature maps are used for predicting bounding boxes at various scales (later layers responsible for larger objects). Thus, the overall idea of SSD is shown in Fig. 8. Some of the activations are passed to the sub-network that acts as a classifier and a localizer.

A. Convolutional Neural Network

CNN is the one of the most famous deep learning approaches where multiple layers are training. It is very successful in computer vision, especially, in an annual competition called the Large-Scale Visual Recognition Challenge (ILSVRC). It is done on a very large database Image net containing 1.2 million images with 1000 classes. Some famous CNN are Lenet, Alex net, Clarifai, SPP, VGG, Google Net, Resnet. A CNN typically comprises multiple convolutional and sub-

ISSN: 2350-0328
International Journal of AdvancedResearch in Science, Engineering and Technology
Vol. 10, Issue 9, September 2023

sampling layers, optionally followed by the fully-connected layers like a standard multi-layer neural network. The main advantage of CNN over fully-connected networks is that they are easier to train and have fewer parameters with the same number of hidden units to identify species. In this paper, we expoilt robustness of a deep convolutional neural network (CNN) through a large dataset with 1000 species. To select the appreciated network, we firstly implement a comparative study on evaluating performances of well-known CNNs such as Alexnet, Caffenet, Googlenet. Alexnet was proposed by Krizhevsky et al. [11] and won ILSVRC 2012. This network has 8 layers and gets 15.3% at top 5 error. Alexnet includes 60 million parameters and 650.000 neurons. The architecture of Alexnet is described in Fig.4. Caffenet presented in [12] is a modified version of Alexnet. It has 5 convolutional layers and 3 fully connected layers. The difference between Caffenet and Alexnet is that some layers are switched to reduce the memory footprint and increase bias-filter value in Caffenet. Googlenet of Szegedy et al. has won ILSVRC 2014. It used a new variant of convolutional neural network called "Inception" for classification. Fig.5 shows a schematic view of Googlenet. It is a very deep neural network model with 22 layers when counting only layers with parameters (or 27 layers if we also count pooling) [8]. The overall number of layers (independent building blocks) used for the construction of the network is about 100. Googlenet incorporates Inception module with the intention of increasing network depth with computational efficiency. Googlenet with 5 million parameters gets 6.7% at top 5 error.

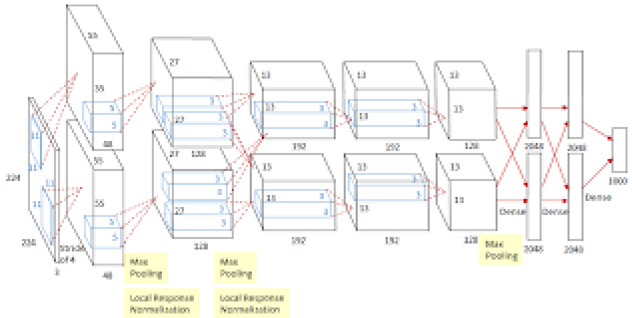


Figure: An illustration of Alexnet's architecture

Figure: A schematic view of Googlenet network

V. EXPERIMENTAL RESULTS

In this system image is taken as input. So first have to upload image using choose file button we can upload an image after uploading image have to click the button. After taking few seconds it will shows the object detection.

ISSN: 2350-0328

International Journal of AdvancedResearch in Science, Engineering and Technology

Vol. 10, Issue 9, September 2023

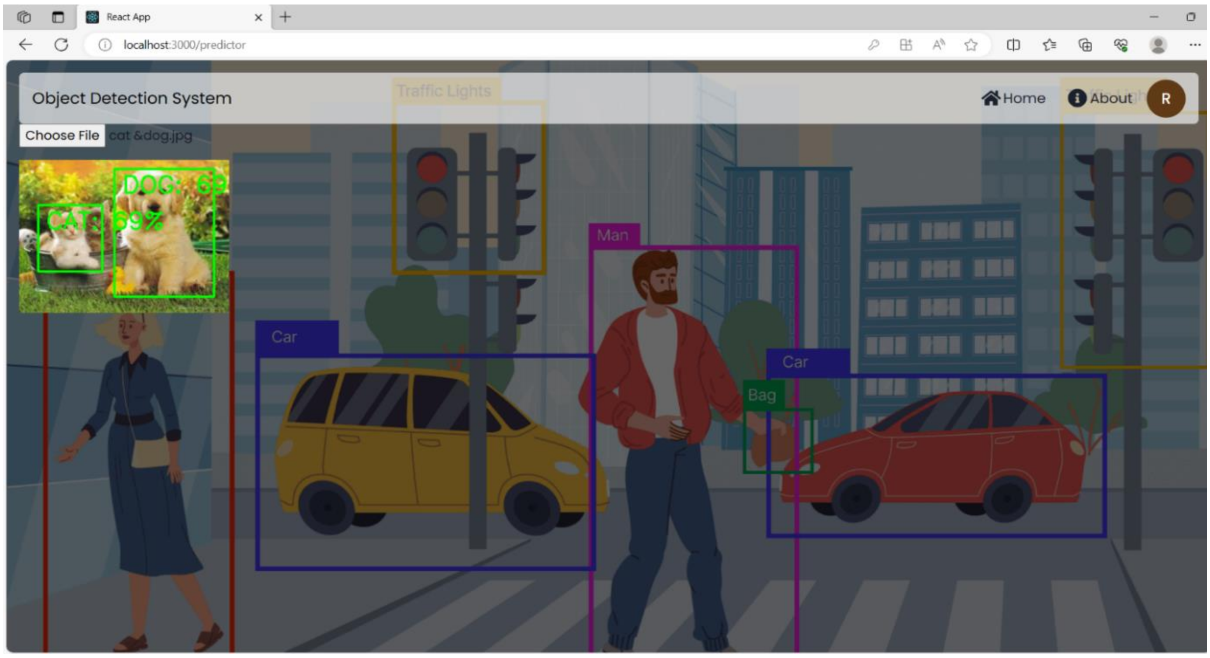


Figure: The user interface displays detected object as systemic

VI. CONCLUSION

An accurate and efficient object detection system has been developed which achieves comparable metrics with the existing state-of-the-art system. This project uses recent techniques in the field of computer vision and deep learning. Custom dataset was created using labelling and the evaluation was consistent. These results come at some computational cost at training time, one needs to train a network per object type and mask type.

By using this thesis and based on experimental results we are able to detect object more precisely and identify the objects individually with exact location of an object in the picture in x, y axis. This paper also provides experimental results on different methods for object detection and identification and compares each method for their efficiencies.

REFERENCES

1.Agarwal, S., Awan, A., and Roth, D. (2004). Learning to detect objects in images viaasparse, part-based representation. IEEE Trans. Pattern Anal. Mach. Intell. 26,1475–1490.doi:10.1109/TPAMI.2004.108

2. Alexe, B., Deselaers, T., and Ferrari, V. (2010). "What is an object?" in Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on (San Francisco, CA: IEEE), 73–80.doi:10.1109/CVPR.2010.5540226

3. Aloimonos, J., Weiss, I., and Bandyopadhyay, A. (1988). Active vision. Int. J. Comput. Vis.1,333–356. doi:10.1007/BF00133571

4. Andreopoulos, A., and Tsotsos, J. K. (2013). 50 years of object recognition: direc-tions forward. Comput. Vis. Image Underst. 117, 827–891. doi: 10.1016/j.cviu.2013.04.005

5. Azizpour, H., and Laptev, I. (2012). "Object detection using strongly-supervised deformable part models," in Computer Vision-ECCV 2012 (Florence: Springer),836–849.

6. Azzopardi, G., and Petkov, N. (2013). Trainable cosfire filters for keypoint detection and pattern recognition. IEEE Trans. Pattern Anal. Mach. Intell. 35, 490–503.doi:10.1109/TPAMI.2012.106

7. Azzopardi, G., and Petkov, N. (2014). Ventral-stream-like shape representation: from pixel intensity values to trainable object-selective cosfire models. Front. Comput. Neurosci. 8: 80.doi:10.3389/fncom.2014.00080

8. Benbouzid, D., Busa-Fekete, R., and Kegl, B. (2012). "Fast classification using sparse decision dags," in Proceedings of the 29th International Conference on Machine Learning (ICML-12), ICML '12, edsJ. Langford and J. Pineau (New York, NY: Omnipress), 951–958.

9. Bengio, Y. (2012). "Deep learning of representations for unsupervised and transfer learning," in ICML Unsupervised and Transfer Learning, Volume 27 of JMLR Proceedings, eds I. Guyon, G. Dror, V. Lemaire, G. W. Taylor, and D. L. Silver (Bellevue: JMLR.Org), 17–36.

10. Bourdev, L. D., Maji, S., Brox, T., and Malik, J. (2010). "Detecting people using mutually consistent pose let activations," in Computer Vision – ECCV2010 – 11th European Conference on Computer Vision, Heraklion, Crete, Greece, September 5-11, 2010, Proceedings, Part VI, Volume6316ofLecture Notes in Computer Science, eds K. Daniilidis, P. Maragos, and N. Paragios (Heraklion: Springer), 168–181.

11. Bourdev, L. D., and Malik, J. (2009). "Poselets: body part detectors trained using 3dhumanposeannotations," in IEEE 12th International Conference on Computer Vision, ICCV2009, Kyoto, Japan, September 27 – October 4, 2009 (Kyoto: IEEE),1365–1372.

12. Cadena, C., Dick, A., and Reid, I. (2015). "A fast, modular scene understanding system using context-aware object detection," in Robotics and Automation (ICRA),2015 IEEE International Conference on (Seattle, WA).
13. Correa, M., Hermosilla, G., Verschae, R., and Ruiz-del-Solar, J. (2012). Human detection and identification by robots using thermal and visual information in domestic environments. J. Intell. Robot Syst. 66, 223–243. doi:10.1007/s10846-011-9612-2
14. Dalal, N., and Triggs, B. (2005). "Histograms of oriented gradients for human detection," in Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, Vol. 1 (San Diego, CA: IEEE), 886–893. doi:10.1109/CVPR.2005.177
15. Erhan, D., Szegedy, C., Toshev, A., and Anguelov, D. (2014). "Scalable object detection using deep neural networks," in Computer Vision and Pattern Recognition Frontiers in RoboticsandAIwww.frontiers November 2015.