REAL TIME SMART OBJECT DETECTION USING MACHINE LEARNING

Madapati Asha Jyothi [1] Mr. M. Kalidas [2]
MCA, IV SEM, Dept. of MCA, Chaitanya Bharathi Institute of Technology (A), Gandipet, Hyderabad-75
Asst.professor, Dept. of MCA, Chaitanya Bharathi Institute of Technology (A), Gandipet, Hyderabad-75

ABSTRACT: Efficient and accurate object detection has been an important topic in the advancement of computer vision systems. With the advent of 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. A major challenge in many of the object detection systems is the dependency on other computer vision techniques for helping the deep learning based approach, which leads to slow and non-optimal performance. in this project, we use a completely 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 data-set, on which a object detection challenge is conducted annually. The resulting system is fast and accurate, thus aiding those applications which require object detection.

Keywords–Deep learning, Machine learning, object detection

1. INTRODUCTION

Object detection is a well-known computer technology connected with computer vision and image processing. With the advent of deep learning techniques, the accuracy for object detection has increased drastically. It focuses on detecting objects or its instances of a certain class (such as humans, flowers, animals) in digital images and videos. There are various applications including face detection, character recognition, and vehicle calculator.

Many problems in computer vision were saturating on their accuracy before a decade. However, with the rise of deep learning techniques, the accuracy of these problems drastically improved. One of the major problems was that of mage classification, which is defined as predicting the class of the mage. A slightly complicated problem is that of mage localization, where the mage contains a single object and the system should predict the class of the location of the object in the mage (a bounding box around the object). The more complicated problem (this project), of object detection involves both classification and localization. in this case, the input to the system will be a mage, and the output will be a bounding box corresponding to all the objects in the mage, along with the class of object in each box.

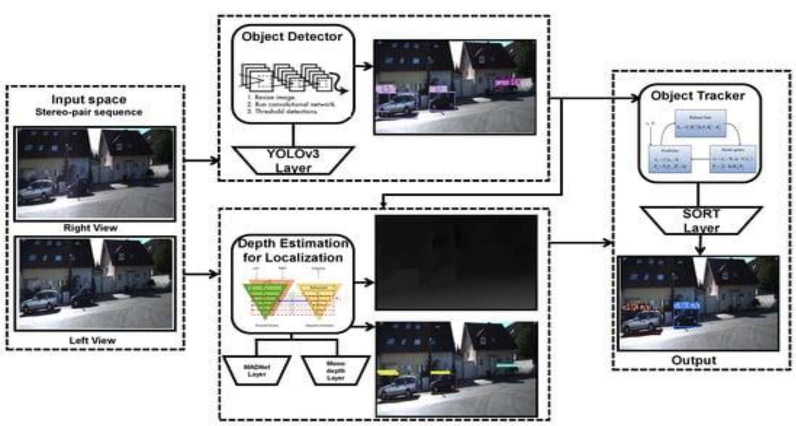

Fig.1: Example figure

Object Detection is the process of finding and recognizing real-world object instances such as car, bike, TV, flowers, and humans out of an images or videos. An object detection technique lets you understand the details of an image or a video as it allows for the recognition, localization, and detection of multiple objects within an image. It is usually utilized in applications like image retrieval, security, surveillance, and advanced driver assistance systems (ADAS).

Object Detection is done through many ways:

- Feature Based Object Detection
- Viola Jones Object Detection
- SVM Classifications with HOG Features
- Deep Learning Object Detection

Object detection from a video in video surveillance applications is the major task these days. Object detection technique is used to identify required objects in video sequences and to cluster pixels of these objects. The detection of an object in video sequence plays a major role in several applications specifically as video surveillance applications. Object detection in a video stream can be done by processes like pre-processing, segmentation, foreground and background extraction, feature extraction. Humans can easily detect and identify objects present in an image. The human visual system is fast and accurate and can perform complex tasks like identifying multiple objects with little conscious thought. With the availability of large amounts of data, faster GPUs, and better algorithms, we can now easily train computers to detect and classify multiple objects within an image with high accuracy.

2. LITERATURE REVIEW

Robust object tracking with online multiple instance learning

In this paper we address the problem of tracking an object in a video given its location in the first frame and no other information. Recently, a class of tracking techniques called "tracking by detection" has been shown to give promising results at real-time speeds. These methods train a discriminative classifier in an online manner to separate the object from the background. This classifier bootstraps itself by using the current tracker state to extract positive and negative examples from the current frame. Slight inaccuracies in the tracker can therefore lead to incorrectly labeled training examples, which degrade the classifier and can cause drift. in this paper we show that using Multiple instance Learning (MIL), instead of traditional supervised learning, avoids these problems and can therefore lead to a more robust tracker with fewer parameter tweaks. We propose a novel online MIL algorithm for object tracking that achieves superior results with real-time performance. We present thorough experimental results (both qualitative and quantitative) on a number of challenging video clips.

A review and comparison of measures for automatic video surveillance systems.

Today's video surveillance systems are increasingly equipped with video content analysis for a great variety of applications. However, reliability and robustness of video content analysis algorithms remain an issue. They have to be measured against ground truth data in order to quantify the performance and advancements of new algorithms. Therefore, a variety of measures have been proposed in the literature, but there has neither been a systematic overview nor an evaluation of measures for specific video analysis tasks yet. This paper provides a systematic review of measures and compares their effectiveness for specific aspects, such as segmentation, tracking, and event detection. Focus is drawn on details like normalization issues, robustness, and representatives. A software framework is introduced for continuously evaluating and documenting the performance of video surveillance systems. Based on many years of experience, a new set of representative measures is proposed as a fundamental part of an evaluation framework.

Handcrafted and Deep Trackers: Recent Visual Object Tracking Approaches and Trends

In recent years visual object tracking has become a very active research area. An increasing number of tracking algorithms are being proposed each year. t is because tracking has wide applications in various real-world problems such as human-computer interaction, autonomous vehicles, robotics, surveillance and security just to name a few. in the current study, we review latest trends and advances in the tracking area and evaluate the robustness of different trackers based on the feature extraction methods. The first part of this work comprises a comprehensive survey of the recently proposed trackers. We broadly categorize trackers into Correlation Filter based Trackers (CFTs) and Non-CFTs. Each category is further classified into various types based on the architecture and the tracking mechanism. in the second part, we experimentally evaluated 24 recent trackers for robustness, and compared handcrafted and deep feature-based trackers. We observe that trackers using deep features performed better, though in some cases a fusion of both increased performance significantly. In order to overcome the drawbacks of the existing benchmarks, a new benchmark Object Tracking and Temple Color (OTTC) has also been proposed and used in the evaluation of different algorithms. We analyze the performance of trackers over eleven different challenges in OTTC, and three other benchmarks. Our study concludes that Discriminative Correlation Filter (DCF) based trackers perform better than the others. Our study also reveals that inclusion of different types of regularization s over DCF often results in boosted tracking performance. Finally, we sum up our study by pointing out some insights and indicating future trends in visual object tracking field

3. METHODOLOGY

There has been a lot of work in object detection using traditional computer vision techniques (sliding windows, deform-able part models). However, they lack the accuracy of deep learning based techniques. Among the deep learning-based techniques, two broad class of methods are prevalent: two stage detection (RCNN, Fast RCNN, Faster RCNN) and unified detection (Yolo, SSD).

Disadvantages:

1. Less prediction rate
2. less accuracy

Here we are proposed YOLOV3 and YOLOV3-TINY models, one of the important fields of Artificial Intelligence is Computer Vision the science of computers and software systems that can recognize and understand images and scenes. Computer Vision is also composed of various aspects such as image recognition, object detection, image generation, image super-resolution and more. Object detection is probably the most profound aspect of computer vision due the number of practical use cases. Object detection refers to the capability of software systems to locate objects in an image/scene and identify each object. It has been widely used for face detection, vehicle detection, pedestrian counting, web images, security systems and driverless cars. There are many ways object detection can be used as well in many fields of practice. Like every other computer technology, a wide range of creative and amazing uses of object detection will definitely come from the efforts of computer programmers and software developers. Getting to use modern object detection methods in applications and systems, as well as building new applications based on these methods is not a straight forward task. Early implementations of object detection involved the use of classical algorithms, the popular computer vision library. However, these classical algorithms could not achieve enough performance to work under different conditions.

Advantages:

1. High accuracy
2. very effective models

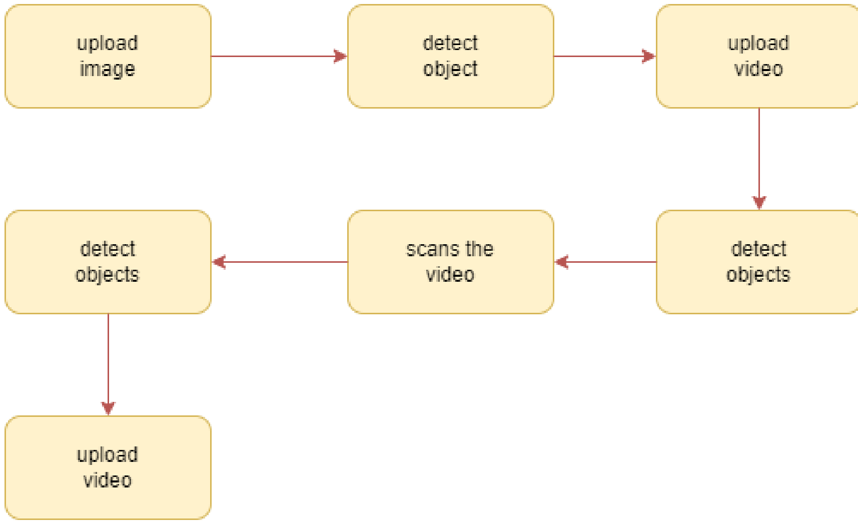

Fig.2: System architecture

YOLOv2 improves the performance by using more anchor boxes and a new bounding box regression method.

YOLOv3 is an enhanced version of the v2 variant with a deeper feature detector network and minor representational changes. YOLOv3 has relatively speedy inference times with it taking roughly 30ms per inference.

YOLOv4 (YOLOv3 upgrade) works by breaking the object detection task into two pieces, regression to identify object positioning via bounding boxes and classification to determine the object's class. YOLO V4 and its successors are technically the product of a different set of researchers than versions 1-3.

MODULES:

1. Data collection:

- **ImageNet:** Image Net data set consists of around 14 million images in total for 21,841 different categories of objects (*data as of 12th Feb 2020*). Some of the popular categories of objects in ImageNet are Animal (fish, bird, mammal, invertebrate), Plant (tree, flower, vegetable) and Activity (sport).
- **Common Objects in Context (COCO):**COCO is a large-scale object detection, segmentation, and captioning dataset. It contains around 330,000 images out of which 200,000 are la-belled for 80 different object categories.
- **Google's Open Images:** Open Images is a dataset of around 9M images annotated with image-level labels, object bounding boxes, object segmentation masks, and visual relationships. It contains a total of 16M bounding boxes for 600 object classes on 1.9M images, making it the largest existing dataset with object location annotations.

2. Data Pre-processing:

- **Read image:** In this step, we store the path to our image dataset into a variable then we created a function to load folders containing images into arrays. But first, we need to import the libraries that we are going to use for this tutorial first.
- **Resize image:** In this step in order to visualize the change, we are going to create two functions to display the images the first being a one to display one image and the second for two images. After that, we then create a function called processing that just receives the images as a parameter.

3. Training & Testing:

Step 1: Annotate some images. During this step, you will find/take pictures and annotate objects' bounding boxes. ...

Step 3: Configuring a Training Pipeline. ...

Step 4: Train the model. ...

Step 5: Exporting and download a Trained model.

4. Modelling:

Given an image or a video stream, an object detection model can identify which of a known set of objects might be present and provide information about their positions within the image.

4. IMPLEMENTATION

4.1. TensorFlow

TensorFlow is a free and open-source software library for dataflow and differentiable programming across a range of tasks. t is a symbolic math library, and is also used for machine learning applications such as neural networks. It is used for both research and production at Google.

TensorFlow was developed by the Google Brain team for Internal Google use. t was released under the Apache 2.0 open-source license on November 9, 2015.

TensorFlow object detection is a computer vision technique that detects, locates, and traces an object from a still image or video. The method allows us to recognize how the models work and provides a fuller understanding of the image or video by detecting objects.

4.2. Fast R-CNN (Region-based Convolutional Neural Network).

Object detection is the process of finding and classifying objects in an image.

One deep learning approach, regions with Convolutional neural networks (R-CNN), combines rectangular region proposals with Convolutional neural network features. R-CNN is a two-stage detection algorithm.

4.3. Single Shot Detector (SSD)

Single Shot detector like YOLO takes only one shot to detect multiple objects present in an image using multibox. It is significantly faster in speed and high-accuracy object detection algorithm. A quick comparison between speed and accuracy of different object detection models on VOC2007.

4.4 YOLO (You Only Look Once)

YOLO is an abbreviation for the term 'You Only Look Once'. This is an algorithm that detects and recognizes various objects in a picture (in real-time). Object detection in YOLO is done as a regression problem and provides the class probabilities of the detected images.

4.5 Region-based Fully Convolutional Network (R-FCN)

To achieve this, R-FCN utilities position-sensitive score maps to address a dilemma between translation-in variance in image classification and translation-variance in object detection.

This R-CNN architecture uses the selective search algorithm that generates approximately 2000 region proposals. These 2000 region proposals are then provided to CNN architecture that computes CNN features. These features are then passed in an SVM model to classify the object present in the region proposal.

4.6 Spatial Pyramid Pooling (SPP-net)

Spatial Pyramid Pooling (SPP) is a pooling layer that removes the fixed-size constraint of the network, i.e. a CNN does not require a fixed-size input image. Specifically, we add an SPP layer on top of the last Convolutional layer. A Pyramid Pooling Module is a module for semantic segmentation which acts as an effective global contextual prior.

5. EXPERIMENTAL RESULTS

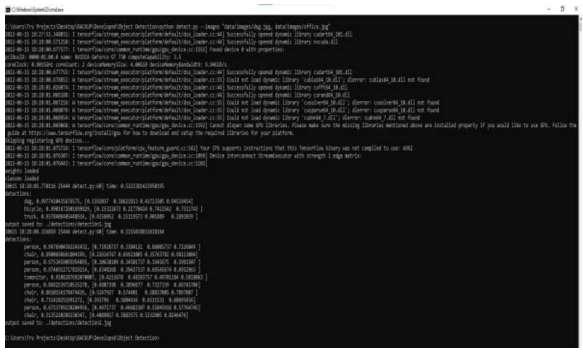

Fig.3: image detection

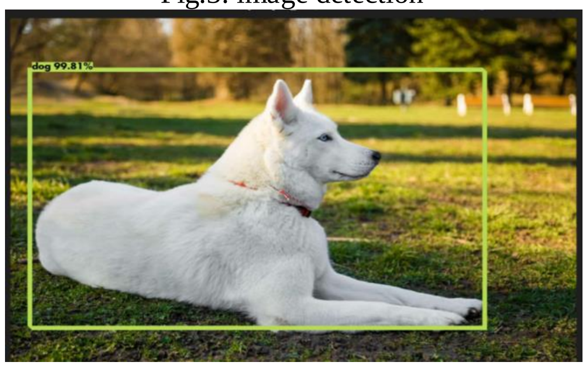

Fig.4: run image

Fig.5: captured data in the image

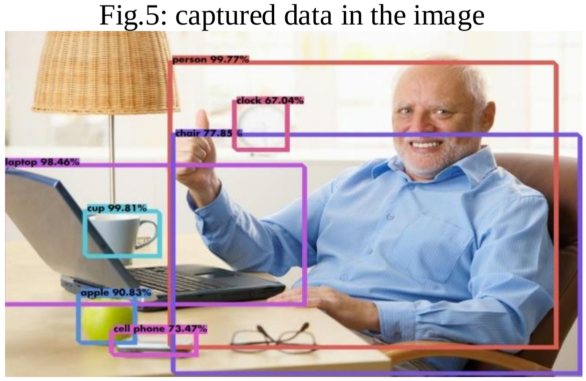

Fig.6: captures accuracy of an image

ISSN: 0972-2750
Vol-12 Issue-11 No.01: 2022

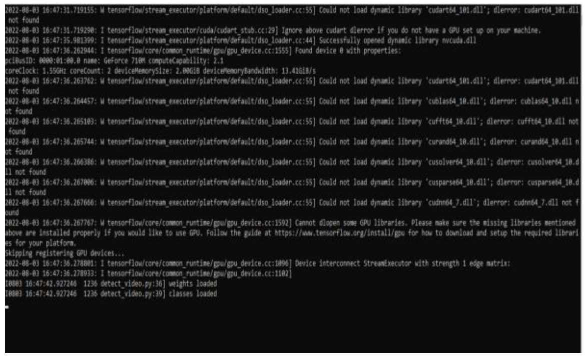

Fig.7: live video

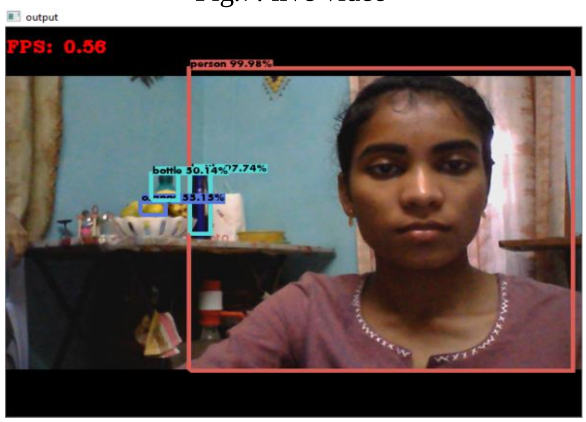

Fig.8: video

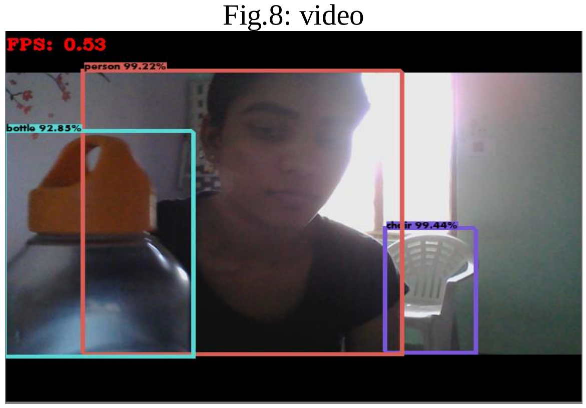

Fig.9: captures accuracy of live video

6. 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 labeling and the evaluation was consistent. This can be used in real-time applications which require object detection for Pre-processing in their pipeline. An important scope would be to train the system on a video sequence for usage in tracking applications. Addition of a temporally consistent network would enable smooth detection and more optimal than per-frame detection.

7. FUTURE WORK

Computer vision is still a developing discipline, it has not been matured to that level where it can be applied directly to real life problems. After few years computer vision and particularly the object detection would not be any more futuristic and will be ubiquitous. For now, we can consider object detection as a sub-branch of machine learning.

REFERENCES

[1] Babenko, B., Yang, M., Belongie, S.: Robust object tracking with online multiple Instance learning. EEE Transactions on Pattern Analysis and Machine Intelligence 33(8), 1619– 1632 (2011) 2

[2] Baumann, A., Boltz, M., Ebling, J., Koenig, M., Loos, H.S., Merkel, M., Niem, W., Warzelhan, J.K., Yu, J.: A review and comparison of measures for automatic video surveillance systems. EURASIP Journal on mage and Video Processing 2008(824726), 1–30 (2008) 2

[3] Bhat, G., Johnander, J., Danelljan, M., Shahbaz Khan, F., Felsberg, M.: Unveiling the power of deep tracking. n: Proceedings of the European Conference on Computer Vision (ECCV), pp. 483–498 (2018) 13

[4] Cehovin, L., Leonardis, A., Kristan, M.: Visual object trac ˇ king performance measures revisited. EEE Transactions on mage Processing 25(3), 1261–1274 (2016) 2

[5] Fiaz, M., Mahmood, A., Javed, S., Jung, S.K.: Handcrafted and deep trackers: Recent visual object tracking approaches and trends. ACM Computing Surveys (CSUR) 52(2), 43 (2019) 1

[6] Henriques, J.F., Caseiro, R., Martins, P., Batista, J.: High-speed tracking with kernelized correlation filters. EEE transactions on pattern analysis and machine ntelligence 37(3), 583–596 (2014) 13

[7] Henriques, J.F., Caseiro, R., Martins, P., Batista, J.: High-speed tracking with kernelized correlation filters. EEE Transactions on Pattern Analysis and Machine ntelligence 37(3), 583–596 (2015) 2

[8] Kasturi, R., Goldgof, D., Soundararajan, P., Manohar, V., Garofolo, J., Bowers, R., Boonstra, M., Korzhova, V., Zhang, J.: Framework for performance evaluation of face, text, and vehicle detection and tracking in video: Data, metrics and protocol. EEE Transactions on Pattern Analysis and Machine ntelligence 31(2), 319–336 (2009)