VISVESVARAYA TECHNOLOGICAL UNIVERSITY

"Jnana Sangama", Belagavi-590018, Karnataka

Technical Seminar Synopsis on

"Fast Yolo:A real time object detection algorithm"

Submitted by

USN	Name
1BI16CS161	Sushmitha M Katti

Under the Guidance of

Dr. Bhargavi M S

Assistant Professor
Department of CSE, BIT
Bengaluru-560004

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

BANGALORE INSTITUTE OF TECHNOLOGY

K.R. Road, V.V.Pura, Bengaluru-560 004

2019-2020

INTRODUCTION

In recent years, object detection has become a significant field of computer vision. The goal of object detection is to detect and classify objects leading to many specialized fields and applications such as face detection and face recognition. Vision is not only the ability to see a picture in ones head but also the ability to understand and infer from the image that is seen. The ability to replicate vision in computers is necessary to progress day to day technology. Object detection addresses this issue by predicting the location of objects through bounding boxes while simultaneously classifying each object in a given image.

LITERATURE REVIEW

A machine learning based intelligent vision system for autonomous object detection and recognition[1] - In this work, authors present an intelligent machine vision system able to learn autonomously individual objects present in real environment. This system relies on salient object detection. In this context authors suggest a novel fast algorithm for visually salient object detection, robust to real-world illumination conditions. Then they used it to extract salient objects which can be efficiently used for training the machine learning-based object detection and recognition unit of the proposed system. Then they provided results of our salient object detection algorithmon MSRA SalientObject Database benchmark comparing its quality with other state-of-the-art approaches.

Real-time object detection and localization with SIFT-based clustering[2] -This paper presents an innovative approach for detecting and localizing duplicate objects in pick-and-place applications under extreme conditions of occlusion, where standard appearance-based approaches are likely to be ineffective. The approach exploits SIFT keypoint extraction and mean shift clustering to partition the correspondences between the object model and the image onto different potential object instances with real-time performance. Then, the hypotheses of the object shape are validated by a projection with a fast Euclidean transform of some delimiting points onto the current image.In order to improve the detection in the case of reflective or transparent objects, multiple object models (of both the same and different faces of the object) are used and fused together. Many measures of efficacy and efficiency are

provided on random disposals of heavily-occluded objects, with a specific focus on real-time processing. Experimental results on different and challenging kinds of objects are reported.

Flip-Invariant SIFT for Copy and Object Detection[3] - This paper proposes a new descriptor, named flip-invariant SIFT (or F-SIFT), that preserves the original properties of SIFT while being tolerant to flips. F-SIFT starts by estimating the dominant curl of a local patch and then geometrically normalizes the patch by flipping before the computation of SIFT. They demonstrate the power of F-SIFT on three tasks: large-scale video copy detection, object recognition, and detection. In copy detection, a framework, which smartly indices the flip properties of F-SIFT for rapid filtering and weak geometric checking, is proposed.

Scalable High Quality Object Detection[4] – This paper demonstrates that learning-based proposal methods can effectively match the performance of hand-engineered methods while allowing for very efficient runtime-quality trade-offs. Using new multi-scale convolutional MultiBox (MSC-MultiBox) approach, they substantially advance the state-of-the-art on the ILSVRC 2014 detection challenge data set, with 0.5 mAP for a single model and 0.52 mAP for an ensemble of two models. MSC-Multibox significantly improves the proposal quality over its predecessor Multibox method: AP increases from 0.42 to 0.53 for the ILSVRC detection challenge. Finally, they demonstrate improved bounding-box recall compared to Multiscale Combinatorial Grouping with less proposals on the Microsoft-COCO data set.

Region-based Convolutional Networks for Accurate Object Detection and Segmentation[5] - In this paper, authors proposed a simple and scalable detection algorithm that improves mean average precision (mAP) by more than 50% relative to the previous best result on VOC 2012—achieving a mAP of 62.4%. Their approach combines two ideas: (1) one can apply high-capacity convolutional networks (CNNs) to bottom-up region proposals in order to localize and segment objects and (2) when labeled training data are scarce, supervised pre-training for an auxiliary task, followed by domain-specific fine-tuning, boosts performance significantly. Since they combined region proposals with CNNs, they call the resulting model an R-CNN or Region-based Convolutional Network.

You Only Look Once: Unified, Real-Time Object Detection[6] - Authors present YOLO, a new approach to object detection. Prior work on object detection repurposes classifiers to

perform detection. Instead, they frame object detection as a regression problem to spatially separated bounding boxes and associated class probabilities. A single neural network predicts bounding boxes and class probabilities directly from full images in one evaluation. Since the whole detection pipeline is a single network, it can be optimized end-to-end directly on detection performance**.**

G-MS2F: GoogLeNet based Multi-Stage Feature Fusion of Deep CNN for Scene Recognition[7] - In this work, the GoogLeNet model is employed and divided into three parts of layers from bottom to top. The output features from each of the three parts are applied for scene recognition, which leads to the proposed GoogLeNet based multi-stage feature fusion (G-MS2F). The product rule is used to generate the final decision for scene recognition from the three outputs corresponding to the three parts of the proposed model. The experimental results demonstrate that the proposed model is superior to a number of state-of-the-art CNN models for scene recognition, and obtains the recognition accuracy of 92.90%, 79.63% and 64.06% on the benchmark scene recognition datasets Scene15, MIT67 and SUN397, respectively.

Face detection using deep learning: An improved faster RCNN approach [8]- In this paper, authors present a new face detection scheme using deep learning and achieve the state-of-the-art detection performance on the well-known FDDB face detection benchmark evaluation. In particular, they improve the state-of-the-art Faster RCNN framework by combining a number of strategies, including feature concatenation, hard negative mining, multi-scale training, model pre-training, and proper calibration of key parameters.

YOLO based Human Action Recognition and Localization[9] - In this paper, authors present an approach to detect, localize and recognize actions of interest in almost real-time from frames obtained by a continuous stream of video data that can be captured from a surveillance camera. The model takes input frames after a specified period and is able to give action label based on a single frame. Combining results over specific time authors predicted the action label for the stream of video. They demonstrate that YOLO is effective method and comparatively fast for recognition and localization in Liris Human Activities dataset.

LIMITATION OF EXISTING SYSTEM

The traditional machine learning methods extract the object features from images and then input the features into a classifier. The traditional methods of extracting features include the histogram of oriented gradient (HOG) method, scale invariant feature transform (SIFT), etc. The methods of classification include the support vector machine (SVM), Bayesian, decision trees, etc. These methods mainly rely on prior knowledge. They are not real-time because they down sample constantly. In addition, these methods have few feature points and edge feature extraction is sometimes not accurate. The core of these methods is feature extraction, and the quality of feature extraction directly affects the method performance. However, in practical applications, these methods are mostly targeted at recognizing specific objects using small datasets and have poor generalization ability. Although machine learning methods are constantly being optimized, from the extraction of low-level features to the emergence of images, the most successful method is the deformable part model (DPM). However, this method has slow detection and depends on the geometric properties of the samples. At present, the traditional machine learning methods cannot meet the efficiency, performance, speed and intelligence requirements of data processing in. OD technology. Among the deep learning methods, region-convolutional neural network(RCNN), Faster region-convolutional neural network(Faster RCNN), You Only Look Once (YOLO) and single shot multibox detector(SSD) are the most widely used methods in OD. However, the current OD methods based on deep learning still have problems due to the slow detection speed and high time consumption.

PROPOSED SYSTEM

In this paper, authors proposed a real-time video OD method, and introduced the Fast YOLO algorithm for OD theory in abstract and then introduce the Fast YOLO framework in detail, including the preprocessing procedure, model training and loss function. Next, they have verified the performance of the Fast YOLO algorithm through some experiments. Considering the theory, proposed methods and experiments, they achieved an excellent performance.

ARCHITECTURE

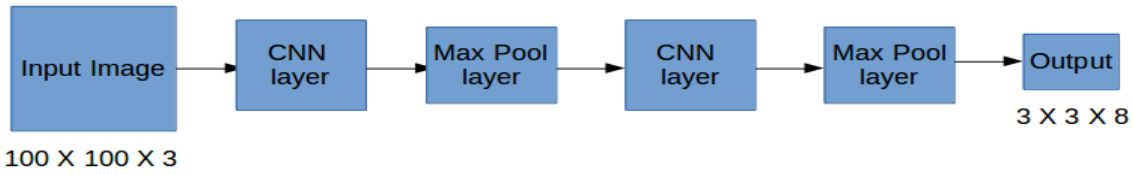

Figure 1. Overall architecture of object detection

METHODOLOGIES/ ALGORITHM

Input: vehicle video parameters: αcoord, αnoobj

Procedure: Obtain each frame of the video

Preprocess the images with the frame difference and background difference methods

For each frame in the video

Divide the image into S∗S grids

Build five different sizes of boxes as bounding boxes for the center of each grid

Compute the probability that the boundary box contains the object

Compute the probability that the center of the object falls within into the grid

Compute the probability of the class for each bounding box

End for

Output: boundary box (x, y. w, h, C), and the probability of classes.

APPLICATIONS

- **Self-Driving Car -** Self-driving cars (also known as autonomous cars) are vehicles that are capable of moving by themselves with little or no human guidance. Now, in order for a car to decide its next step, i.e. either to move forward or to apply breaks, or

to turn, it must know the location of all the objects around it. Using Object Detection techniques, the car can detect objects like other cars, pedestrians, traffic signals, etc.

- **Face Detection and Face Recognition -**Face detection and recognition are perhaps the most widely used applications of computer vision. Every time you upload a picture on Facebook, Instagram or Google Photos, it automatically detects the people in the images.
- **Action Recognition -** The aim is to identify the activity or the actions of one or more series of images. Object Detection is the core concept behind this which detects the activity and then recognizes the action.
- **Object Counting -** We can use Object Detection algorithms for counting the number of objects in an image or even in real-time videos. Counting the number of objects is helpful in a variety of ways, including analyzing the performance of a store, or estimating the number of people in a crowd.

REFERENCES

1. Ramík, Dominik Maximilián, Christophe Sabourin, Ramon Moreno, and Kurosh Madani. "A machine learning based intelligent vision system for autonomous object detection and recognition." Applied intelligence, vol. 40, no. 2, pages:358-375 2014.
2. Piccinini, Paolo, Andrea Prati, and Rita Cucchiara. "Real-time object detection and localization with SIFT-based clustering." Image and Vision Computing, vol. 30, no.8, pages: 573-587 2014.
3. Zhao, Wan-Lei, and Chong-Wah Ngo. "Flip-invariant SIFT for copy and object detection." IEEE Transactions on Image Processing, vol. 22, no. 3, pages: 980-991 2012.
4. Szegedy, Christian, Scott Reed, Dumitru Erhan, Dragomir Anguelov, and Sergey Ioffe. "Scalable, high-quality object detection." arXiv preprint arXiv, pages:1412-1441 2014.
5. Girshick, Ross, Jeff Donahue, Trevor Darrell, and Jitendra Malik. "Region-based convolutional networks for accurate object detection and segmentation." IEEE transactions on pattern analysis and machine intelligence vol. 38, no. 1, pages: 142-158 2015.

6. Redmon, Joseph, Santosh Divvala, Ross Girshick, and Ali Farhadi. "You only look once: Unified, real-time object detection." In Proceedings of the IEEE conference on computer vision and pattern recognition, pages: 779-788 2016.
7. Tang, Pengjie, Hanli Wang, and Sam Kwong. "G-MS2F: GoogLeNet based multi-stage feature fusion of deep CNN for scene recognition." Neurocomputing vol. 225, pages: 188-197 2017.
8. Sun, Xudong, Pengcheng Wu, and Steven CH Hoi. "Face detection using deep learning: An improved faster RCNN approach." Neurocomputing vol. 299 pages: 42-50 2018
9. Shinde, Shubham, Ashwin Kothari, and Vikram Gupta. "YOLO based Human Action Recognition and Localization." Procedia computer science vol. 133, pages: 831-838 2018.
10. Lu, Shengyu, Beizhan Wang, Hongji Wang, Lihao Chen, Ma Linjian, and Xiaoyan Zhang. "A real-time object detection algorithm for video." Computers & Electrical Engineering vol. 77, pages: 398-408 2019.