International Journal of Multidisciplinary Engineering in Current Research - IJMEC

Volume 10, Issue 5, May-2025, http://ijmec.com/, ISSN: 2456-4265

Object Detection and Tracking using Deep Learning and Artificial Intelligence for Video Surveillance Applications

Guttula Kalyan
PG scholar, Department of MCA, CDNR collage, Bhimavaram, Andhra Pradesh.
B.S.Murthy
(Assistant Professor), Master of Computer Applications, DNR collage, Bhimavaram, Andhra Pradesh.

Abstract: ***In the evolving landscape of intelligent surveillance systems, object detection and tracking play a critical role in enhancing situational awareness and automated decision-making. This research presents an efficient deep learning-based framework for real-time object detection and tracking within video surveillance environments, leveraging the strengths of Convolutional Neural Networks (CNN) and You Only Look Once version 3 (YOLOv3). The system is trained on urban vehicle datasets and evaluated using KITTI and COCO datasets, enabling both single and multiple object detection. Key performance metrics such as accuracy, precision, confusion matrix, and mean Average Precision (mAP) are utilized for validation. Detected objects are further tracked across consecutive frames using Simple Online Real-time Tracking (SORT), enabling trajectory analysis. The framework's robustness in varying lighting and environmental conditions makes it suitable for applications in traffic density estimation, autonomous vehicle navigation, and smart city infrastructure. Comparative analysis with existing models underscores the proposed method's capability in achieving accurate, fast, and real-time tracking, highlighting its potential for deployment in modern intelligent transport systems.***

Keywords : Object Detection , CNN , python , surveillance, real time tracking

INTRODUCTION

Over the past years domains like image analysis and video analysis has gained a wide scope of applications. CV and AI are two main technologies dominating technical society. Technologies try to depict the biology of human. Human vision is the sense through which a perception of outer 3D world is perceived. Human Intelligence is trained over years to distinguish and process scene captured by eyes. These intuitions acts as a crux to budding new technologies. Rich resource is now accelerating researchers to excavate more details form the images. These developments are due to state-of the-art methods like CNN. Applications from Google, Facebook, Microsoft, and Snapchat are all results of tremendous improvement in Computer vision and Deep learning. During time, the vision-based technology has transformed from just a sensing modality to intelligent computing systems which can understand the real world. Computer vision applications like vehicle navigation, surveillance and autonomous robot navigation find Object detection and tracking as important challenges. For tracking vehicles and other real word objects, video surveillance is a dynamic environment. In this paper, efficient algorithm is designed for object detection and tracking for video Surveillance in complex environment.

Object detection and tracking goes hand in hand for computer vision applications. Object detection is identifying object or locating the instance of interest in-group of suspected frames. Object tracking is identifying trajectory or path; object takes in the concurrent frames. Image obtained from dataset is, collection of frames. Basic block diagram of object detection and tracking is shown in Fig. 1. Data set is divided into two parts. 80 % of images in dataset are used for training and 20 % for testing. Image is considered to find objects in it by using algorithms CNN and YOLOv3. A bounding box is formed across object with Intersection over union (IoU) > 0.5. Detected bounding box is sent as references for neural networks aiding them to perform Tracking. Bounded box is tracked in concurrent frames using Multi Object Tracking (MOT). Importance of this research work is used to estimate traffic density in traffic junctions, in autonomous vehicles to detect

various kinds of objects with varying illumination, smart city development and intelligent transport systems.

Objective of the project:

Data is the new oil in current technological society. The impact of efficient data has changed benchmarks of performance in terms of speed and accuracy. The enhancement is visualizable because the processing of data is performed by two buzzwords in industry called Computer Vision (CV) and Artificial Intelligence (AI). Two technologies have empowered major tasks such as object detection and tracking for traffic vigilance systems. As the features in image increases demand for efficient algorithm to excavate hidden features increases. Convolution Neural Network (CNN) model is designed for urban vehicle dataset for single object detection and YOLOv3 for multiple object detection on KITTI and COCO dataset. Model performance is analyzed, evaluated and tabulated using performance metrics such as True Positive (TP), True Negative (TN), False Positive (FP), False Negative (FN), Accuracy, Precision, confusion matrix and mean Average Precession (mAP). Objects are tracked across the frames using YOLOv3 and Simple Online Real Time Tracking (SORT) on traffic surveillance video. This paper upholds the uniqueness of the state of the art networks like DarkNet. The efficient detection and tracking on urban vehicle dataset is witnessed. The algorithms give real-time, accurate, precise identifications suitable for real-time traffic applications**.**

LITERATURE SURVEY

Learning Framework for Robust Obstacle Detection, Recognition, and Tracking

This paper introduces a general framework for detection, recognition, and tracking preceding vehicles and pedestrians based on a deep learning approach. The proposed framework combines a novel deep learning approach with the use of multiple sources of local patterns and depth information to yield robust on-road vehicle and pedestrian detection, recognition, and tracking. The proposed system is first based on robust obstacle detection to identify obstacles appearing along the road that are likely to be vehicles and pedestrians, implemented as an efficient adaptive U-V disparity algorithm. Second, the results from the obstacle detection stage are input into a novel vehicle and pedestrian recognition system based on a deep learning model that processes multiple sources of depth information and local patterns. Finally, the results from the recognition stage are used to track detected vehicles or pedestrians in the next frame by means of a proposed tracking and validation model. The proposed framework has been thoroughly evaluated by inputting several vehicle and pedestrian data sets that were collected under various driving conditions. Experimental results show that this framework provides robust vehicle and pedestrian detection, recognition, and tracking with high accuracy, and also satisfies the real-time requirements of driver assistance systems.

Detecting Abnormal Events in University Areas

This paper presents a distinct video surveillance system which took place in the Lebanese International University Saida- Campus., which is considered as a very crowded environment., and reveals if there is an unusual event. Our main target is to apply simple procedures that will be present as a future"s benchmark. The work is split into three major parts., starting by dividing the video frame into zones., then to compute the magnitude of optical flow in each., and finally to analyze these data and classify it., based on a logical threshold., as normal or abnormal events. We implement our results based on Histogram of Magnitudes for each zone (HOM) and the outcome met our expectations.

Detection of unwanted traffic congestion based on existing surveillance system using in freeway via a CNN-architecture trafficnet

Detection of traffic congestion is important for route guidance using in intelligent transport system (ITS) to prevent jam escalation. Although the surveillance

system has been used in freeway for years, it is hard to automatically identify and report traffic congestion in complicated transportation scene according to various illumination, weather and other disturbances. The detection process based on human eye is time-consuming and tedious as the machine detection accuracy is not high enough to meet the requirements of practical applications. In this paper, a new classifier is proposed using convolutional neural networks (CNN) to generate four TrafficNet based on two championships of ILSVRC including AlexNet and VGGNet. Instead of using fully-connected layers in AlexNet and VGGNet, a support vector machine (SVM) are used after CNN architecture. Congestion and non-congestion images are trained and tested through this new structure. Image database with more than 30000 images are extracted from existing traffic surveillance video and corresponding labels are added manually. With database, those TrafficNet are trained and tested, detection accuracy and training time of those TrafficNet are compared. The experimental results show that the accuracy of proposed method can reach up to 90%, which is much higher than traditional method based on feature extraction without deep learning.

The Research of Target Tracking Algorithm Based on an Improved PCANet

The feature extraction method will greatly affect the performance of the target tracking algorithm. In traditional feature extraction methods, feature descriptors are manually designed, such as HOG features, etc. which does not express spatial information very well. In this paper, a novel target tracking method is proposed. This method uses a lightweight deep learning model, called PCANet network, to extract features. In the previous work, the number of PCA layer filters is determined through a large number of experiments. In this paper, the number of filters in the PCA layer is determined by the cumulative contribution rate, and which achieves the adaptive adjustment of the network parameters. Firstly, the region of interest is acquired by particle filtering; Secondly, the depth characteristics of the image are extracted via PCANet; Finally, the target is determined by the SVM classifier. The results of the experiment show that this algorithm has strong robustness, because the target tracking can be tracked accurately under the condition of illumination change, occlusion and rapid movement.

Real-Time Detection, Tracking and Classification of Multiple Moving Objects in UAV Videos

Unnamed Aerial Vehicles (UAVs) are becoming increasingly popular and widely used for surveillance and reconnaissance. There are some recent studies regarding moving object detection, tracking, and classification from UAV videos. A unifying study, which also extends the application scope of such previous works and provides real-time results, is absent from the literature. This paper aims to fill this gap by presenting a framework that can robustly detect, track and classify multiple moving objects in real-time, using commercially available UAV systems and a common laptop computer. The framework can additionally deliver practical information about the detected objects, such as their coordinates and velocities. The performance of the proposed framework, which surpasses human capabilities for moving object detection, is reported and discussed.

Object Tracking with Shallow Convolution Feature

Traditional target tracking algorithm uses manual extraction of features, which is difficult to cope with the challenges of rotation and occlusion and deformation. Based on the deep learning method, the convolution neural network is used to extract the features. Because of lost a lot of spatial information in the convolution process, it's easy to make the tracking target drift. In this paper, we use a shallow convolutional network without second training to extract features for tracking, which combines hard negative mining technology and bounding box regression to refine the target location. We have compared our tracker performance with others state-of-the-art tracker. The obtained experimental results in OTB dataset demonstrate the effectiveness of our proposed tracker has outperformed the compared tracking algorithms.

Convolutional Neural Networks Based Fire Detection in Surveillance Videos

The recent advances in embedded processing have enabled the vision based systems to detect fire during surveillance using convolutional neural networks (CNNs). However, such methods generally need more computational time and memory, restricting its implementation in surveillance networks. In this research paper, we propose a cost-effective fire detection CNN architecture for surveillance videos. The model is inspired from GoogleNet architecture, considering its reasonable computational complexity and suitability for the intended problem compared to other computationally expensive networks such as AlexNet. To balance the efficiency and accuracy, the model is fine-tuned considering the nature of the target problem and fire data. Experimental results on benchmark fire datasets reveal the effectiveness of the proposed framework and validate its suitability for fire detection in CCTV surveillance systems compared to state-of-the-art methods.

Cell Tracking with Deep Learning and the Viterbi Algorithm

We present a cell tracking pipeline that combines deep cell segmentation with a Viterbi algorithm tracker to accurately detect and track cells in microscopy videos. Our pipeline handles large illumination shifts, large appearance variability in the cells, and heavy occlusion from other cells and debris. We first train a Fully Convolutional Network (FCN) to detect the cells, then track the cells across frames using a tracker based on the Viterbi algorithm. We evaluate our algorithm on a dataset featuring Escherichia coli (E. coli) where the experimental goal is to immobilize the E. coli using blue light, thus making the dataset especially challenging due to large illumination shifts. Our results demonstrate that despite these challenges, our pipeline is able to accurately detect and track the cells.

An object tracking method using deep learning and adaptive particle filter for night fusion image

In this paper, we propose an online visual tracking algorithm for fused sequences via deep learning and adaptive Particle filter (PF). Our algorithm pretrains a simplified Convolution Neural Network (CNN) to obtain a generic target representation. The outputs from the hidden layers of the network help to form the tracking model for an online PF. During tracking, the moving information guides the distribution of particle samples. The tests illustrate competitive performance compared to the state-of-art tracking algorithms especially when the target or camera moves quickly.

Online Multi-Object Tracking Using Selective Deep Appearance Matching

In this paper, we focus on designing appearance matching network and solving computational bottleneck problem of it. From the development of deep neural network and graphic device(GPU), many research topics in computer vision (e.g. detection, classification) achieved state-of-the-art performance using convolutional neural network(CNN). In multi-object tracking, also, there have been several works which used CNN for extracting appearance feature of targets. Although, deep appearance feature improved an accuracy of tracking, it increased processing time and made an algorithm hard to be applied in real-world situation. So, we propose a simple technique to improve speed by removing redundant appearance matchings. Also, we propose a structure of joint-input siamese network and method to train it. We verify the performance of our work by comparison with recent online trackers.

III. Existing System

Object tracking is a very challenging task in the presence of variability Illumination condition, background motion, complex object shape partial and full object occlusions. Object detection and location in digital images has become one of the most important applications for industries to ease user, save time and to achieve parallelism. This is not a new technique but improvement in object detection is still required in order to achieve the targeted objective more efficiently and accurately. The main aim of studying and researching computer vision is to

simulate the behavior and manner of human eyes directly by using a computer and later on develop a system that reduces human efforts.

Disadvantage

1. Objects are not easily Tracking & Detected.
2. Less Accuracy.
3. Human effort is more.
4. More Time Tacking process.

Proposed System

In order to overcome the issue of detection, tracking related to object movement and appearance. Most of the algorithm focuses on the tracking algorithm to smoothen the video sequence. In this project using python and OPENCV module we are detecting objects from videos and webcam. This application consists of two modules such as 'Browse System Videos' and 'Start Webcam Video Tracking'.

Browse System Videos Using this module application allow user to upload any video from his system and application will connect to that video and start playing it, while playing if application detect any object then it will mark that object with bounding boxes, while playing video if user wants to stop tracking then he need to press 'q' key from keyboard to stop video playing.

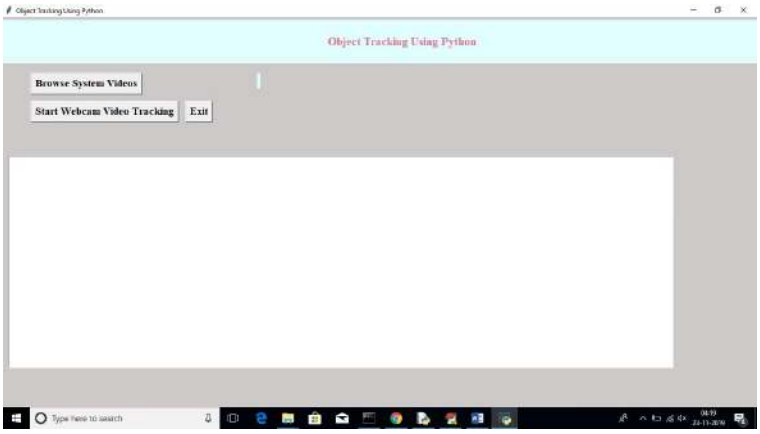
Start Webcam Video Tracking: Using this module application connect itself with inbuilt system webcam and start video streaming, while streaming if application detect any object then it will surround that object with bounding boxes, while playing press 'q' to stop web cam streaming.

Advantage

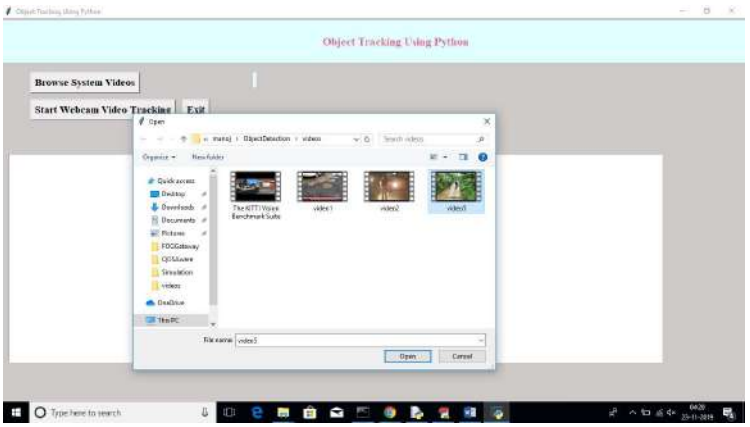
1. Objects are Tracking & Detected Easily.

Results Analysis

Double click on 'run.bat' file to get below screen

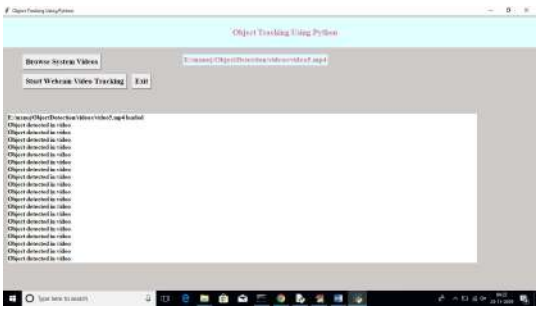


Now click on 'Browse System Videos' button to upload videos from system

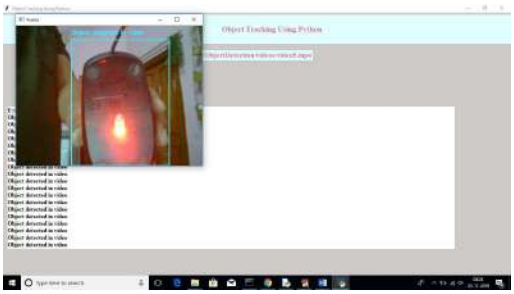


In above screen I am uploading one video, after upload will get below screen

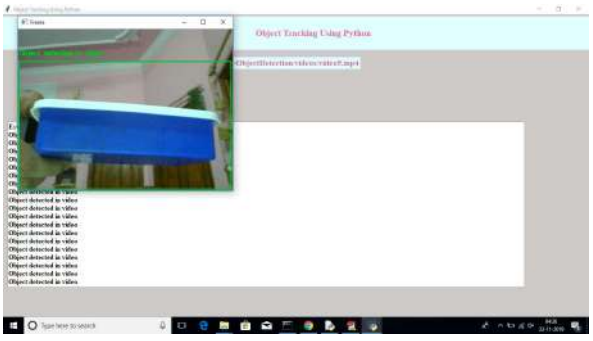
In above video we can see application start tracking objects from video and mark them with bounding boxes. Similarly we can upload any video and track objects from video



In above screen now click on another button called 'Start Webcam Video Tracking' to connect application to web cam and start streaming. After connecting to webcam will get below screen



In above screen we can see objects is getting tracked from webcam also. In above screen it track computer mouse from web cam video



CONCLUSION

The inclusion of Artificial Intelligence to solve Computer vision tasks has outperformed the image processing approaches of handling the tasks. The CNN model trained to on road vehicle dataset for single object detection, achieved a validation accuracy of 95.7 % for auto, 95.5% for car and 96 % for heavy vehicles for day images. The high validation accuracy is because of huge amount of data on which it is trained from each class. Performance metrics are tabulated for day, evening and NIR images. Multiple object detection is implemented using YOLOv3 for KITTI and COCO dataset. Performance metrics is tabulated for YOLOv3 on considered classes of images. Higher the precession value of class greater will be mAP value. The mAP value depends on image chosen for calculation. IoU of 0.5 is ideal for detection and tracking. mAP values can be enhanced by increasing true positive values. Results of performance metrics is totally dependent on image data set used. Further objects are detected in video based on region of interest. The performance measures measured such as speed and color of vehicle, type of vehicle, direction of vehicle movement and the number of vehicles in ROI. Multiple object tracking is implemented for traffic surveillance video using YOLOv3 and OpenCV. Multiple objects are detected and tracked on different frames of a video. Further training the models on powerful GPUs and by increasing the number of images evaluate the models on other datasets and modify the design if required to make the model more robust and suitable for real-time applications.

REFERENCES

[1] V. D. Nguyen et all., "Learning Framework for Robust Obstacle Detection, Recognition, and Tracking", IEEE Transactions on Intelligent Transportation Systems, vol. 18, no. 6, pp. 1633-1646, June 2017.

[2] Zahraa Kain et all, "Detecting Abnormal Events in University Areas", 2018 International conference on Computer and Applications(ICCA),pp. 260-264, 2018.

[3] P. Wang et all., "Detection of unwanted traffic congestion based on existing surveillance system using in freeway via a CNN-architecture trafficnet", IEEE Conference on Industrial Electronics and Applications (ICIEA), Wuhan, 2018, pp. 1134-1139.

[4] Q. Mu, Y. Wei, Y. Liu and Z. Li, "The Research of Target Tracking Algorithm Based on an Improved PCANet", 10th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC), Hangzhou, 2018, pp. 195-199.

[5] H. C. Baykara et all., "Real-Time Detection, Tracking and Classification of Multiple Moving Objects in UAV Videos", 29th IEEE International Conference on Tools with Artificial Intelligence (ICTAI), Boston, MA, 2017, pp. 945-950.

[6] W. Wang, M. Shi and W. Li, "Object Tracking with Shallow Convolution Feature", 9th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC), Hangzhou, 2017, pp. 97-100.

[7] K. Muhammad et all., "Convolutional Neural Networks Based Fire Detection in Surveillance Videos", IEEE Access, vol. 6, pp. 18174-18183, 2018.

[8] D. E. Hernandez et all., "Cell Tracking with Deep Learning and the Viterbi Algorithm", International Conference on Manipulation, Automation and Robotics at Small Scales (MARSS), Nagoya, 2018, pp. 1-6.

[9] X. Qian et all., "An object tracking method using deep learning and adaptive particle filter for night fusion image", 2017 International Conference on Progress in Informatics and Computing (PIC), Nanjing, 2017, pp. 138-142.

[10] Y. Yoon et all., "Online Multi-Object Tracking Using Selective Deep Appearance Matching", IEEE International Conference on Consumer Electronics - Asia (ICCE-Asia), Jeju, 2018, pp. 206-212.

[11] H. S. Bharadwaj, S. Biswas and K. R. Ramakrishnan. "A large scale dataset for classification of vehicles in urban traffic scenes", Proceedings of the Tenth Indian Conference on Computer Vision, Graphics and Image Processing,ACM,2016.

[12] Mohana et.al., "Performance Evaluattion of background modeling methods for object Detection and Tracking",International Conference on Inventive systems and Control (ICISC).

[13] G. Chandan et.al., "Real Time Object Detection and Tracking Using Deep Learning and OpenCV", International Conference on Inventive Research in Computing Applications (ICIRCA), 2018.

[14] Mohana et.al., "Elegant and efficient algorithms for real time object detection, counting and classification for video surveillance applications from single fixed camera" International Conference on Circuits, Controls, Communications and Computing (I4C),2016.

[15] Mohana et.al., "Simulation of Object Detection Algorithms for Video Survillance Applications", 2nd International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud),2018.