RESEARCH ARTICLE OPEN ACCESS

Object Tracking Using Deep Reinforcement Learning

Mr. Anil P Jawalkar[1], Sai Sangavi M[2], Sneha M[3], Meghana Rao M[4]

1 Assistant Professor, Department of Information Technology,
Malla Reddy Engineering College for Women (UGC-Autonomous),
Hyderabad, India
Email:anil.jawalkar022@gmail.com

2 Department of Information Technology,
Malla Reddy Engineering College for Women(UGC-Autonomous),
Hyderabad, India
Email:sangavireddy04@gmail.com

3 Department of Information Technology,
Malla Reddy Engineering College for Women(UGC- Autonomous),
Hyderabad, India
Email : snehamotlakunta6@gmail.com

4 Department of Information Technology,
Malla Reddy Engineering College for Women(UGC- Autonomous),
Hyderabad, India
Email : meghanarao242@gmail.com

Abstract:

In this research, we offer an effective visual tracker that, through sequential actions honed using deep neural networks, directly captures a bounding box enclosing the target object in a video. Pre-trained using several training video sequences, the proposed deep neural network to govern tracking actions is then modified while actually tracking to allow for online response to a change in target and background. Deep reinforcement learning (RL) and supervised learning are both used for the pre-training. Even partially labelled data can be successfully used for semi-supervised learning when RL is applied. The suggested tracker is validated to attain a competitive performance at three times the speed of existing deep network-based trackers by the analysis of the object tracking benchmark data set.

I. INTRODUCTION

One of the main issues in the science of computer vision is how to discover a bounding box tightly containing the target moving object in every frame of a movie. This is the goal of visual object tracking. Visual tracking algorithms have made great strides in recent years, but there are still many difficult problems that arise from a variety of tracking impediments, including motion blur, occlusion, changing lighting, and background clutter. This category includes the YOLO approach. We won't pick the image's interesting sections for this. Instead, we use a single algorithm run to forecast the classes and bounding boxes of the entire image and a single neural network to recognize multiple objects. When compared to other classification methods, the YOLO algorithm is quick. Our algorithm processes 45 frames per second in real time.

Convolutional neural networks (CNNs) tracking techniques have recently been presented for robust tracking and greatly enhanced tracking performance with the aid of rich feature representation by deep hidden layers. Pre-trained CNNs, which are used in the early efforts, are trained using extensive classification data sets like ImageNet. The challenge of adaptation to object shape deformation and illumination changes in tracking, however, cannot be solved by a CNN pre-trained on a classification data

set because the deep CNN is not suitable for online adaptation. In this research, we attempt to create a reinforcement learning (RL) strategy to efficiently use the partially labelled video sequences, in contrast to the existing deep CNN-based trackers employing SL.

The proposed action-driven deep tracker pursues a target change through repetitive actions controlled by an action-decision network (ADNet) built with deep CNN architecture.

We recast the visual tracking problem as selecting sequential actions and propose an RL-based training method. The ADNet is intended to generate actions in order to determine the location and size of the target object in a new frame.

The ADNet learns the policy that chooses the best actions to track the target based on its current position. The policy network in the ADNet is designed with a CNN, with the input being an image patch cropped at the previous state's. There are fewer searching steps in this action selection process than in the sliding window or candidate sampling approaches. Furthermore, because our method precisely localizes the target by selecting actions, post processing such as bounding box regression is not required. We also propose a deep RL learning algorithm for training the ADNet.

The entire training framework is divided into two stages: SL and RL. By simulating tracking on training sequences, we produce training samples for RL. Using the rewards obtained from the tracking simulation, the network is trained with deep RLusing a policy gradient. The proposed system successfully learns the unlabeled frames even in the semi-supervised (SS) condition, when the training frames are only partially labelled, by allocating the rewards in accordance with the outcomes of the tracking simulation.

II. EXISTING SYSTEM

Before a decade, many computer vision problems had reached saturation in terms of accuracy. However, with the rise of deep learning techniques, the accuracy of these problems has dramatically improved.

The first works make use of pre-trained CNNs that have been trained on a large-scale classification data set, such as ImageNet.

However, because deep CNNs are not suitable for online adaptation, a CNN pre-trained on a classification data set is insufficient to solve the problem of adaptation to object shape deformation and illumination changes in tracking. This CNN-based tracking-by-detection approach achieves a breakthrough in tracking performance, but it suffers from an inefficient exhaustive search strategy that explores the region of interest and selects the best candidate based on network scores.

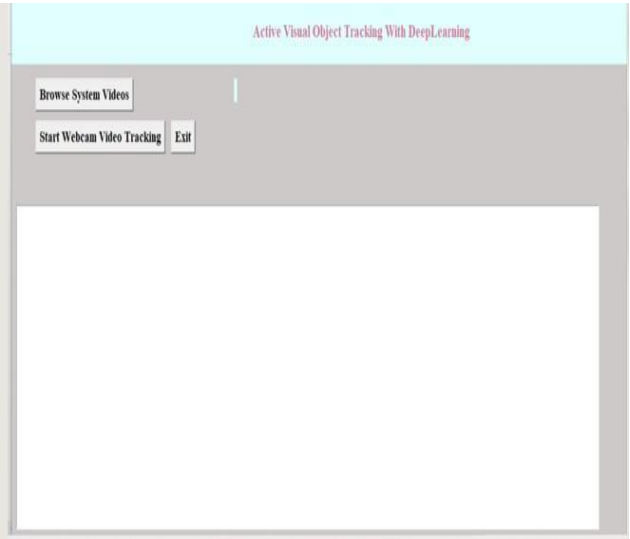
III . PROPOSED SYSTEM

These These approaches, known as dense optical flow, aid in estimating the motion vectors of each pixel in a video frame.

Sparse optical flow methods, such as the Kanade-Lucas-Tomashi (KLT) feature tracker, locate a small number of feature points in an image.

Kalman Filtering is a well-liked signal processing approach that forecasts the location of a moving surveillance item based on previous motion data. The guidance of missiles was one of this algorithm'searly uses.

IV. RESULTS



V. CONCLUSION

In this publication, we propose a novel action-driven method for visual tracking based on deep convolutional networks. The proposed tracker is monitored by an ADNet, which iteratively seeks a target object through sequential actions. The action-driven tracking proved popular significantly to the reduction of tracking computation complexity. Furthermore, RL allows for the use of partially labelled data, which can significantly contribute to the creation of training data with minimal effort. The proposed tracker achieves state-of-the-art performance in 3 frames/s, which is three timesfaster than existing deep network-based trackers using a tracking-by-detection strategy, according to the evaluation results.

Furthermore, the short model of the proposed tracker achieves an actual-time speed (15 frames/s) through adjusting the meta parameters of the ADNet, with an accuracy that outperforms state- of-the-artwork actual time trackers. A correct and green surveillance item detection machine has been advanced which achieves similar metrics with the present latest machine. This mission makes use of current strategies with inside the discipline of pc imaginative and prescient and deep learning.

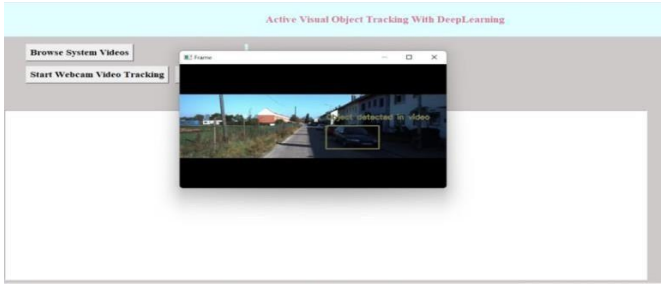


Fig. 1 In above we can see application start tracking objects from video and mark them with bounding boxes.

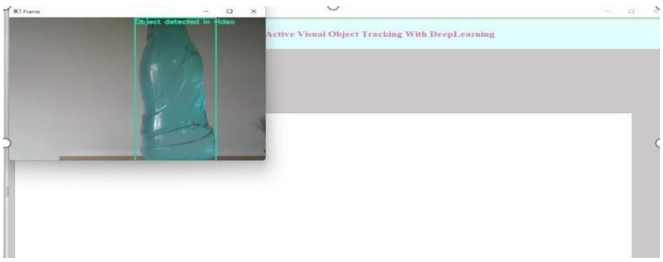


Fig. 2 In above we can see object is getting tracked from web cam also. In above screen it tracks water bottle from web cam video.

REFERENCES

[1] N. Wang and D.-Y. Yeung, "learning a deep compact image representation for visual tracking," in proc. Adv. Neural inf. Process. Syst., 2013, pp. 809–817.

[2] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, "high-speed tracking with kernelized correlation filters," IEEE trans. Pattern anal. Mach. Intell., Vol. 37, no. 3, pp. 583–596, mar. 2015.

[3] H. Grabner, C. Leistner, and H. Bischof, "Semi-supervised on-line boosting for robust tracking," in Proc. Eur. Conf. Comput. Vis., 2008, pp. 234–247.

[4] B. Babenko, M.-H. Yang, and S. Belongie, "Robust object tracking with online multiple instance learning,"

[5] M. Danelljan, G. Häger, F. Khan, and M. Felsberg, "Accurate scale estimation for robust visual tracking," in Proc. Brit. Mach. Vis. Conf., Nottingham, U.K., Sep. 2014.

[6] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, "High-speed tracking with kernelized correlation filters," IEEE Trans. Pattern Anal. Mach. Intell., vol. 37, no. 3, pp. 583–596, Mar. 2015.