Deep Learning on Multi-Object Tracking

By

SAHIL BANSAL

Abstract

Multiple Object Tracking (MOT) is the process of localizing multiple moving objects over time in a video sequence. MOT has a wide range of applications namely, human-computer interaction, surveillance and security, video communication and compression, augmented reality, traffic control, medical imaging, and video editing. Generally, MOT is a two-step process that includes object detection and association. Initially, a unique identifier is assigned to every detected object in the first frame and then motion trajectories of the detected objects were extracted. All the objects are detected and their track is kept in each subsequent frame of a video sequence. Afterward, the trajectories of each detected object are determined in the current frame based on its location in the previous frame. MOT aims to determine better object association to increase the affinity between objects in the subsequent frames. However, accurate multiple object tracking is very challenging. The challenges are either due to object deformation namely, pose variation, occlusion, and background clutter, or due to the dynamic environmental variations namely, fog, snow, rain, and dust particles. In order to address these challenges, a lot of work has been proposed exploiting deep learning. In this chapter, we will review the work proposed in this direction and will exhaustively compare them based on the recent performance metrics.

Introduction

Multi-Object Tracking or MOT is a technique used to study how sensors or any visual systems track multiple moving objects which was developed by Zenon pylyshyn. He did so for studying sustained visual attention in a dynamic environment in 1988. It was invented for supporting visual indexing theory(FINST theory).In this Theory, he used the analogy of fingers as indexes. After this a ton of experiments were conducted for further understanding human's visual and cognitive function.

MOT relies on detecting and identifying targets within videos where targets may be , for instance, pedestrians, vehicles or animals regardless of their appearance or quantities. Its main aim is identifying objects in a video and tracking them as a set of trajectories with good accuracy.

1.1 Uses and types of Object Tracking

This is used for a variety of use cases involving different types of input footage, Where input can be an image or a video, or a real time cctv recording, or any live video, all impacts the algo used for creating object tracking applications.

Here, we are going to describe a few popular uses and types of object tracking such as visual tracking, image tracking and video tracking.

1.1.1 Visual Tracking

This is a good research topic in computer vision that is applied in a large range of everydays scenarios. Its main goal is to estimate the future position of a visual target that was initialized without availability of the rest of the video. For example - in latest cars while reversing back camera shows the path where our car going is supposed to come in a direction.

1.1.2 Video Tracking

In video tracking, moving objects were located within videos. Hence, video tracking objects are capable of processing real time or pre recorded footage. Often, there's an indication around the object being tracked, for example, a surrounding square that follows the object, showing the user how the object is being tracked.

Different applications play an important role in video analytics, in scene understanding for security, military, transportation and other industries.Today, a wide range of computer vision and deep learning applications uses video tracking technology.

1.1.3 Image Tracking

Image Tracking is the continuous tracking of images as they move in the setting. It is ideal for datasets with highly contrasting images, few patterns, asymmetry and multi-identifiable diff. between image set and image of interest.

Image Tracking relies on computer vision to argue and detect images after that image is predetermined.

1.1.4 Object Tracking Camera

New Object tracking methods are used in real-time video stream of any camera, so the video feed of any USB camera or an IP camera can be used. By individual frames to a tracking algorithm, we can perform object tracking easily.

What makes object tracking difficult ?

Main Challenges in object tracking do start from the object that's the image. Image makes it difficult for object tracking to effectively perform detections on an image.

Therefore, we will discuss more issues with the task of object tracking and methods of preventing or dealing with these issues in short.

1. Training and tracking speed -

    These methods are not only supposed to accurately perform detections and localize objects of interest but also to do this in a lesser possible amount of time. It is especially imperative for real-time objects tracking models.

    To decrease time taken for a model, we have to choose the right method for this or have to customize methods. Like fast R-CNN and faster R-CNN can be used to increase speed of that most common R-CNN method. Design choices besides the detection framework also influence the balance between speed and accuracy of a model.

2. Background Distractions -

Background of an image is the main factor impacting the accuracy of an image. A blurry Background or a single colour background makes it easier for an AI system to detect an object while a busy background or same colour as object or for a light coloured object it would be harder to detect small objects or even for large objects.

3. Multiple Spatial Scales -

That Object to be tracked can be in different sizes and aspect ratios. These ratios can confuse our algorithm into believing objects are scaled larger or smaller than their actual size which will impact detections and detection speed.

In order to deal with these problems, we can implement different methods such as feature maps, anchor boxes, image pyramids and feature pyramids.

4. Occlusion -

In AI vision tasks using deep learning, Occlusion happens when multiple come too close to each other so that they even seem to merge or they even get merged sometimes. This causes many problems with object tracking because the occluded objects are seen as one and objects will be tracked incorrectly. The system gets confused and identifies the initially tracked object as a new object.

Occlusion sensitivity can be used to prevent this misidentification by allowing user to understand which parts of the image are most important for an object tracking system to classify. It basically refers to a small measure of the network's sensitivity to occlusion in different data regions. It is done using small subsets of original datasets.

Related Work :-

Authors & year	Methods	Summary
Ohay Angah, Albert Y. Chen (2020)	Gradient based method with rematching based on MOT accuracy	This paper is based on tracking multiple construction workers through deep learning and challenges like mismatch and mis detection due to occlusion and identity switches makes this harder. They aimed to improve performance of current multiple workers tracking in three stages i.e., detection, matching and re-matching. They proposed two improvements in human tracking - gradient based method for location prediction and re-matching.
Gioele Ciaparrone, Francisco Luque Sanchez, Siham Tabik, ´ Luigi Troiano, Roberto Tagliaferri, Francisco Herrera (2019)	Faster R-CNN, SSD, CNNs, Correlation filters, kalman filters,etc. many were explained.	They presented all MOT based algorithms employing deep learning techniques by focusing single-camera videos and 2-D data in brief.They characterized four steps for MOT - detection, feature extraction, affinity computation, association.Approach to association algorithm were explained and a numerical comparison on results of MOT Challenge were explained datasets were also provided.

Haoxiang Liang , Huansheng Song , Huaiyu Li , and Zhe Dai (2020)	You Look Only Once(YOLO) version 3 and KCF	In this paper, they used DL and MOT for making a vehicle counting system. Firstly, a vehicle dataset from a high surveillance camera is constructed and a vehicle detection model is obtained using YOLO. Secondly, an improved method based on KCF is proposed to avoid single feature and single scale defects. Finally an automatic and accurate counting system is designed. This algorithm has practical application in counting vehicles in complex highway scenes.
Tao Yang, Cindy Cappelle, Yassine Ruichek and Mohammed El Bagdouri (2019)	Discriminant Correlation filter based deep learning tracker	In this paper,DCF based CNN is used in deep learning for MOT. Individual tracker is used to estimate the position for each object.Two different pre-trained methods are used as feature extractors. They analyzed and tested their report on MOT2D2015 and MOT17 benchmarks for MOT.

Bassam A. Y. Alqaralleh , Sachi Nandan Mohanty , Deepak Gupta , Ashish Khanna , K. Shankar , And Thavavel Vaiyapuri	Cluster-based RNN-T tracking method	This paper proposed an effective method for animal tracking. It incorporates two mains places - fuzzy logic based clustering algorithm and RNN-T based algorithm for animal tracking. This method comes in handy as compared to other methods for animal tracking and many other factors like noise, camera movements, position,etc. Also be considered by some improvements.
Noor M. Al-Shakarji, Filiz Bunyak, Guna Seetharaman, Kannappan Palaniappan (2017)	Robust detection based method	This paper proposed a MOT system that saves time with good accuracy. Robust detection method allies on an efficient but discriminative target description and a two-step combined local & global data association scheme.

Samuel Scheidegger, Joachim Benjaminsson, Emil Rosenberg, Amrit Krishnan, Karl Granstrom	DNN based method with PMBM filters	This paper deals with the problem of monocular cameras used in automotive industries that make observations in 2D planes and can't measure distance to targets.They used DNN to detect and measure distance between single input images. When PMBM filters were used with DNN detectors, they achieved 3D Tracking using only mono- camera images. This method comes under one of the top performer KITTI object tracking benchmarks with 20 frames per second.
Jerome Berclaz, François Fleuret, Engin Tu¨retken, and Pascal Fua (2011)	Linear Programming with k-shortest path algorithms	This paper shows formalizing the motions of target as flows along the edges of the graph of spatio- temporal locations and hence reduced a difficult problem to Linear programming. And by relying on the k-shortest path algorithm for optimization of our problem, they reduce the complexity to a small fraction from the LP problem.

Keni Bernardin and Rainer Stiefelhagen	Clear MOT Metrics for evaluation of MOT systems.	This paper has proposed two metrics for evaluation of MOT. The MOTA and MOTP are applicable to a wide range of tracking tasks and objective comparison of different tracking systems. The results show that proposed metrics indeed reflect the strengths and weaknesses of the various used systems and can be easily compared to overall performance and many scenarios.
Dina Chahyati, Mohamad Ivan Fanany, Aniati Murni Arymurthy (2017)	Faster R-CNN	In this paper, Faster R-CNN methods are used for detection and two methods were compared for object association. Into results it was seen that simple minimum Euclidean distance association performs well compared to SNN methods because SNN suffered much error in higher number of ID and occlusion.
Changjiang Yang, Ramani Duraiswami and Larry Davis (2005)	Hierarchical Particle Filter	In this paper, an efficient and robust particle filter based algorithm is used for object tracking. This method characterizes tracked objects using colour and edge orientation features. The resulting algorithm maintains multiple hypotheses and offers robustness against occlusions.

Yingkun Xu, Xiaolong Zhou, Shengyong Chen, Fenfen Li	Comparison between different tracking methods.	In this paper, the author summarised and analyzed different tracking methods of deep learning in MOT which are top ranked in benchmark tests. Firstly, they have investigated functionality and classified these methods into categories. Author did experimental comparisons between them and also analyzed their limitations and got some useful conclusions.
Hao Jiang, Sidney Fels and James J. Little (2007)	Linear Programming	In this paper, Linear programming is used for MOT where the inter object interaction matrix is convex and intra object term quantifying object state continuity may use any metric. This method has lesser average complexity and is able to find the global optimum with higher probability.
Xin Li, Kejun Wang, Wei Wang and Yang Li (2010)	Kalman Filter	In this paper, Three Conditions that are object occlusion, occlusion by object background, splits and merges, are mainly responsible for unsatisfying tracking results. The algorithm proposed is valid for human and vehicle image sequence. This method seems to be effective in unsatisfying conditions.

Wenhan Luo, Junliang Xing (2021)	Multiple Object tracking evaluation	In this paper, Different MOT evaluation metrics are presented.It has presented a unified problem formulation and several ways of categorization of existing methods. And also described the main components of MOT and discussed evaluation of MOT algorithms including public datasets, metrics , benchmark tests, etc.
Thierry Chateau, Najoua Essoukri Ben Amara (2019)	Spatio - Temporal interlaced encoding video and DCNN detector	Spatio-temporal variation of objects between images are converted into interlaced images. A specialized detector is used for detecting these images and a classical association algorithm is used for association between objects.This showed an interlacing video can be much helpful in MOT systems and have many advantages.
Zenon Pylyshyn (2014)	Experiments were performed	In this paper, Four different experiments were performed on Multiple object tracking systems. Occlusion is the main challenge in tracking of objects. Their studies suggest that observers are at better tracking of independently moving identical objects.

Alexander Scheel, Christina Knill, Stephan Reuter, and Klaus Dietmayer	Multi-sensor tracking of vehicles using high resolution radars	In this paper, a MOT approach for tracking of vehicles is implemented by high resolution radars. SMC implementation of the LMB filter for extended objects avoids preprocessing steps. Porting the filter to real-time capable languages can make this future proof also.
Jun Xiang, Guoshuai Zhang, And Jianhua Hou	Feature Representation and Bayesian Filtering	Most challenging part of any MOT project is to design a robust affinity model for data association. To resolve the detector's limitations, they designed a RNN based Bayesian filtering for estimating targets.

Deep Learning based MOT

In computer vision, Object tracking in videos is most important. Some of their applications are :- autonomous driving, video surveillance, human computer interface, etc. The main objective of MOT is to track objects by keeping their identities in order to draw their trajectories. It is not easy as it seems many challenges like frequent occlusion, interaction among targets and so on can be a big problem to a MOT system. One of the method i.e, Convolutional neural network achieved remarkable success in pattern recognition, image processing

Table of Contents:

1. Introduction
2. Related work
3. Deep learning based MOT
3.1 CNN based algorithm
3.2 Fusion based algorithm
4. Performance metrics for algorithm evaluation
4.1 Description of performance metrics
4.2 Comparison of deep learning based algorithm
5. Recent advances in MOT
6. Conclusion

Sara Bouraya Jr. and Abdessamad Belangour "Multi object tracking: a survey", Proc. SPIE 11878, Thirteenth International Conference on Digital Image Processing (ICDIP 2021), 118780I (30 June 2021); https://doi.org/10.1117/12.2602901

Vidushi meel “What is Object Tracking? – An Introduction” vlog on viso.ai website (July 10, 2021); https://viso.ai/deep-learning/object-tracking/

Qi Chu , Wanli Ouyang , Hongsheng Li , Xiaogang Wang , Bin Liu , Nenghai Yu, University of Science and Technology of China, University of Sydney Department of Electronic Engineering, The Chinese University of Hong Kong kuki@mail.ustc.edu.cn.; https://openaccess.thecvf.com/content_ICCV_2017/papers/Chu_Online_Multi-Object_Tracking_ICCV_2017_paper.pdf