Traffic Management using IoT and Deep Learning Techniques: A Literature Survey

Aakifah Hassan[1], Abriti Chakraborty[2], Aishwarya Suresh[3], Dhimanth Shukla[4], Asst. Prof. Anupama Girish[5]

Dayananda Sagar College of Engineering, Bangalore, Affiliated to VTU

Abstract- The primary goal of this study is to alleviate traffic congestion created by today's ineffective traffic control system. An automatic traffic management system is proposed for vehicle identification and counting, as well as automatic traffic signal timing, because vehicle flow detection is a key aspect of today's traffic management system. The processing engine receives video input from the cameras. In the streaming video of these routes at an intersection, the values will be read frame by frame. The camera transfers all of the collected input images to the neural network-based processing engine. The traffic flow displays the current traffic situation over a set time interval and aids with traffic management and control, particularly when there is heavy traffic. It will prioritize emergency vehicles such as ambulances and fire trucks.

Keywords: Yolo, Deepsort, Counting, Traffic

Introduction

In this day and age, when innovation has beyond all limitations, it has become much easier to resolve general human concerns, one of which is traffic congestion. Gridlock has risen quickly in recent years, resulting in undesirable consequences such as road rage, accidents, air pollution, and, most importantly, fuel waste. Unwise traffic in the board structures is one of the major causes of traffic congestion. The first gas-lit traffic signal was built in London in the 1860s to control traffic caused by neighboring horse carriages, and it was physically maintained by cops. Since then, traffic signals have been changed to allow for the smooth flow of traffic. The electric traffic signal was introduced in the mid-nineteenth century, and it was quickly replaced by mechanical traffic signals, which are still in use in many urban areas today. This system works as expected, with the lights changing at regular intervals, but it wasn't long before people realized there was a flaw in the system. Many times, automobiles were held up for no reason because the signal would be red in any case, even though the other street was empty.

Related Work

In this section, some papers in the field of object detection and tracking and traffic management are reviewed. This survey starts by going through some of the older algorithms for object detection and then moving on to the most recent and efficient one.

For object detection, Ross Girshick[1] presented a Fast Region-based Convolutional Network technique (Fast R-CNN). R-CNN is sluggish since it conducts a ConvNet forward pass for each item suggestion. Training is a single stage in the rapid R-CNN architecture. It accepts a whole image as well as a series of object suggestions as input. The network then creates a feature map by processing it via numerous convolutional and max pooling layers. A region of interest (RoI) pooling layer is then used to convert the feature map into a fixed-length feature vector. The author ends by stating that there may be strategies that have yet to be identified that will allow dense boxes to perform as well as sparse proposals.

Yuhao Xu and Jiakui Wang[2] in their paper have used Faster RCNN as their base. To detect objects, Faster RCNN first uses the Region Proposal Network (RPN) to generate a set of candidate bounding boxes. An additional branch for tracking is being proposed. They use the track branch to extract track features from the RoI feature vector, then use this feature to calculate the distance between different vehicles.If it is the same vehicle then distance should be less and if it is different vehicles distance should be large. They apply the widely used triplet loss to optimize track branch. In conclusion, some of the advantages of the proposed method are reduction in the amount of calculations and making full use of multi-loss during training and inference. Their proposed method could give them a result of 57.79% mAP and high-performance vehicle tracking.

Anima Pramanik et al[3], have created two novel object detection and tracking models, granulated RCNN (G-RCNN) and multi-class deep SORT (MCD-SORT), respectively. The G-RCNN is an advanced version of Faster RCNN which has two parts. The first portion is the foreground region proposal network (FRPN), which is a granulated deep CNN that delivers foreground RoIs across the video frame and the second part deals with classification of objects in RoI. The use of DeepSORT results in an increase of search space and reduces object tracking speed. All the assignments between the target object and existing track-lets of the same class are computed and solves the above mentioned problem. The authors have achieved a mAP of 80.6% and also reduced runtime while increasing accuracy.

Tuan Linh Dang et al [4], presented a new object tracking architecture that is an upgraded version of the deep sort yolov3 architecture. This helped to overcome the issues of the original method by adding two contributions - identity switches, in case the YOLO object detection misses the object and no detected bounding boxes are forwarded to the DeepSORT components, these objects are detected in subsequent frames and assigned a new object ID using the Dlib tracker and second being operating speed, any detected object from YOLO is sent immediately to the DeepSORT detection. Hence, object detection and object tracking are conducted in parallel.

Aderonke A. Oni et. al [5] focused on creating a vehicle counting framework that is to be used on metropolitan streets in Nigeria, explicitly to be introduced on common roads and overhead bridges. The goal was to study the existing vehicle counting frameworks in the country and design a more stable and enhanced system. After comparing algorithms, YOLO and Discriminative Correlation Filter with Channel and Spatial Reliability were chosen for the proposed system.It displays a screen with a visualization of the entire procedure and records the count data. Ideally, the system would take input from a camera mounted by a road and transmit the vehicle count to other systems for further processing or archiving it for later analysis. The significance of this system incorporates assessing traffic streams on a given road and understanding traffic patterns and the factors that can affect them, and optimization of existing manual traffic management systems.

Urban mobility has quickly become one of the most important aspects of smart city development in India. There have been a slew of studies published in the last few years based on the adoption of a smart traffic management system to combat the traditional preset time span framework, which is responsible for the great majority of the undesired blockage in rush hour traffic jams. The key difference between most previous models is the type of framework and sensors utilized to calculate the thickness of traffic in a given corridor. They all want to overcome the limitations of the traditional framework by combining multiple sensors and algorithms to create a more intelligent system.Currently, the traffic structure is not dependent on traffic density, and each street is allotted a specific amount of preset time. This causes gridlock as a result of long red-light delays and timings given for roadways in a city that should fluctuate throughout peak on-off hours but in reality do not. These lights have predetermined signal timing delays and do not adapt to changing traffic density. When traffic density exceeds a certain threshold on one side, a longer green light time is required to improve traffic flow. Naga Harsha J et. al[6], proposed a system in this paper that uses Ultrasound sensors along with Image processing that works on a Raspberry Pi platform, calculates the vehicle density and dynamically allots time for different levels of traffic. This thus permits better signal control and viable administration of traffic consequently diminishing the likelihood of a crash. By utilizing Internet Of Things(IoT), ongoing information from the framework can be gathered, stored and managed on a cloud. This information can be used to decipher the signal term on the off chance that any of the detecting gear comes up short, and furthermore for future examination.

Mr. Nikhil Chhadikar et. al[7] have proposed a system for classifying an object as a specific type of vehicle. Haar Cascade Classifier is used to detect a car and count the number of passing vehicles on the specific road using traffic videos as input. Viola Jones Algorithm is used for training these cascade classifiers. It is then modified to find unique objects in the video, by tracking each car in a selected region of interest. This is one of the quickest methods for successfully identifying, tracking, and counting an automobile object, with a 78 percent accuracy. The scale factor value affects the detection rate of this system, with different scale factor values offering variable detection rates. The scale factor value that gives the classifier the best performance should be found in order to get a high detection rate. According to the scientists, developing a skilled and reliable vehicle recognition system will be a difficult task in the future.

Alex Bewley et al[8],proposed a basic framework for identifying data linkage between frames using Kalman filtering in picture space and the Hungarian technique. Although the process appears to be simple, the end result is high frame rates. It is based on the use of a detection-based tracking framework for MOT (Multiple Object Tracking). The study examines a practical solution to multiple object tracking, with the primary goal of effectively associating objects for online and real-time applications. To this purpose, detection quality has been discovered as a crucial influencing element in tracking performance, with altering the detector improving tracking by up to 18.9%This solution provides tracking accuracy comparable to state-of-the-art online trackers while just utilizing a simple combination of already existing techniques such as the Kalman Filter and Hungarian algorithm for the tracking constituent. The tracker updates at a rate of 260 Hz, which is around 20 times faster than other state-of-the-art trackers due to the convenience of the tracking method.

Nicolai Wojke et al.[9] proposed a practical solution to multiple object tracking that emphasized simple, effective algorithms. They use appearance information to increase SORT's performance in this paper. They can now track objects through longer periods of not being visible or partially covered as a result of this enhancement, substantially reducing the amount of identification shifts. They put much of the complex algorithm used to calculate into an offline pre-training stage, where they employ a deep association measure on a very big human re-identification dataset, in the spirit of the original approach. Using nearest neighbor queries in the visual appearance space, they construct measurement-to-track linkages during the online application process. Extensions lower the number of identity shifts by 45 percent in experiments, resulting in overall competitive performance at high frame rates.They were also able to obtain cutting-edge performance while being exceedingly fast.

Xiao-jun Tan et. al[10] discuss the vital strategy of video-based vehicle discovery that has a place with an exemplary issue of movement division. Background subtraction or background learning is one of the most widely utilized technologies for recognising moving objects (vehicles). The approaches are divided into two categories: frame-oriented and pixel-oriented. In frame-oriented approaches, a preset limit is used to determine whether the visual scene is moving. The present edge is deducted as the backdrop if the variations of the current casing and its archetype are not exactly the limit. These approaches are simple to use and consume little CPU. They've been successful in locating intruders in interior settings, but they're not practical in rush hour congestion scenes since the light isn't consistent outside; also, fluctuations in brightness levels are interpreted as movement. A pixel-oriented technique, on the other hand, obtains the backdrop by calculating the average value of each pixel over a period far longer than the time it takes for moving objects to traverse the field of vision.

Joseph Redmon et al[11], presented a complete and new method for object detection at the time by Classifiers having been remodeled to do detection in previous work on object detection. Instead, they took object detection to be a regression problem with spatially separated areas of interest and related probabilities of the classes.

In a single evaluation, only one neural network predicts areas of interest and class probabilities straight from the pictures.Since the whole detection pipeline is a single network, it can be made optimum for the detection performance by utilizing the pipeline from the beginning to the end .YOLO does not perform as well as state-of-the-art detection systems when it comes to object localization, but it is much less likely to predict false positives. To conclude, it learns very broad object representations. On both the Picasso Dataset and the People-Art Dataset, it outperforms all the other detection algorithms at the time, including DPM and R-CNN, by a considerable margin when generalizing from normal photos to art works.Their integrated model design is very fast and their version processes frames at 45 frames per second in real time. Fast YOLO, a smaller variant of the same network, processes 155 frames per second while also maintaining twice the mAP of other real-time detectors.

Bo Yang and Ram Nevatia[12] put forward an online learning method for multitarget tracking. Here, the detection response instead of only focusing on producing selective movement and image for all models instead associates into tracklets of various levels to produce the final tracks. The tracking problem consists of an online CRF model and further is adapted for minimizing energy. This approach is more potent towards spatially close targets with identical appearances while dealing with camera motion. The effective algorithm proposed here is in association with low cost energy. The algorithm identifies each target, motion and appearance are implemented to deliver a discriminative descriptor. The descriptors are based upon the speed and distance between pairs of tracklets, meanwhile the appearance descriptors are based on color histograms in order to differentiate targets.This approach yields better on most evaluation scores. Mostly tracked score improves by about 10%, further the fragments reduced by 17%.The recall increases by 2%, precision by 4%.

Thanh-Nghi Doan and Minh-Tuyen Truong[13] propose a robust model that combines both YOLOv4 and Deepsort. This new model is able to determine objects with improved accuracy and rapid computation time by implementing simple and effective algorithms. The algorithm combines the detection and tracking process. The detection is achieved using background subtraction and tracking is done using the kalman filter. A realistic video database is used which includes the commonly seen vehicles. The proposed approach and implementation yields better results by exceeding the original by 11% AP and 12% of AP50 for nearly all field scenarios of the dataset at a real time speed of ¬32 fps.

Tianming Yu et al[14] proposed an unsupervised and brief method based on the features learned from deep CNN in order to improve the traditional background subtraction method. The traditional method is opted here as the deep methods used in the advancement of the background subtraction are supervised. The supervised methods have high computational costs and only work in certain scenes. Meanwhile, the traditional background subtraction methods are of minor costs of computation and are applicable to most general scenes. The proposed method generates much more definite front view object detection results. This is achieved by designing the fundamental features. From the lower layer of the pretrained CNN, the low level features of the input images are extracted. The main features are reserved to promote the dynamic background model. Results show that the proposed implementation fundamentally improves the performance of conventional background subtraction methods.

Junliang Xin et al[15] implements variable object tracking methods in two stages. This works for heavy visual surveillance situations with only a singular camera. The set of reliable tracklets are generated in the primary stage, while In the secondary stage, the detection feedbacks are collected from a transitory sliding window to deal with uncertainty caused due to occlusions of the object in order to produce a collection of possible tracklets. Furthermore, in the transitory window, they are associated by the Hungarian algorithm on a dual modified tracklets association cost matrix to acquire the final optimum association. The resulting proposed implementation fundamentally enhances tracking veracity and efficiency in heavy observable surveillance environments with various occlusions.

Outline of the survey:

The table below provides a basic summary of the papers reviewed including the methodology and the respective conclusions and results derived in each.

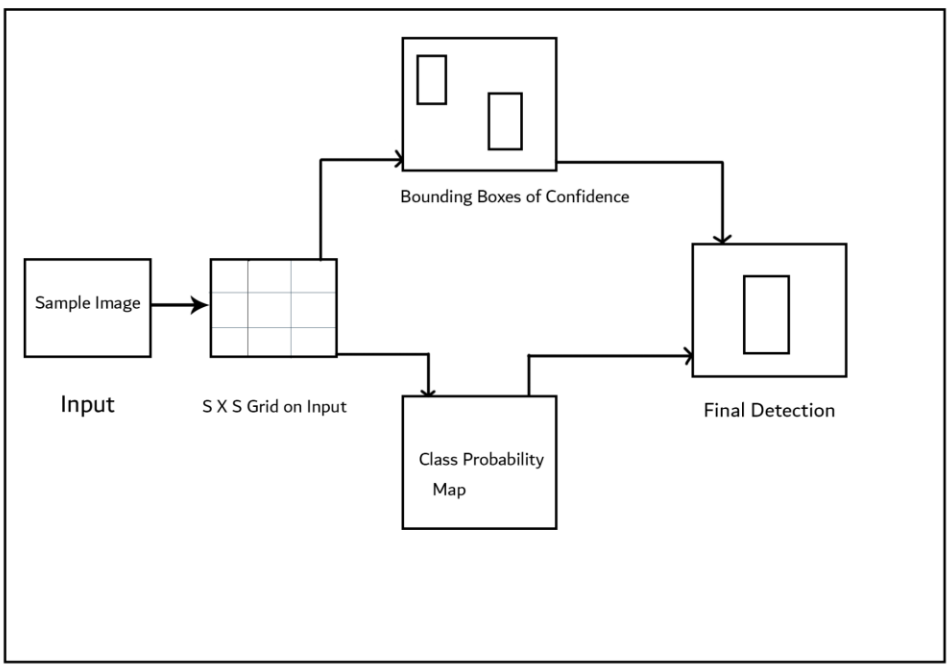
Year	Paper number	Methodology	Conclusion and Result
2015	[1]	In this architecture the training is done in a single stage and so it is much faster compared to the standard R-CNN.	It has a superior detection quality than R-CNN, SSPnet, improved training and testing speed and also accuracy.
2019	[2]	Building upon the Fast R-CNN methodology, an additional branch for tracking is introduced along with the triplet loss method.	It has reduced the amount of computation and can achieve 57.79% mAP and high performance in vehicle detection.
2021	[3]	Introduction of the concept of Granulation into deep CNN architecture and using MCD-SORT.	It signifies the importance of granulation technique in RoI map generation and how it helps in better and accurate object detection.
2020	[4]	The implementation of Dlib tracker along with the YOLO and Deep SORT architecture resulted in a better object tracking methodology.	It reduces the number of identity switches and also obtains higher FPS.

2019	[5]	Based on the YOLO detection method, this research built a vehicle counting system. The vehicle counting system's tracking module employs the DCF-CSR tracking algorithm.	The main programme is made up of three separate modules that work together to keep the system running.
2018	[6]	It has three ultrasound sensors that detect traffic density and a camera that monitors traffic on the route. These devices are linked to a Raspberry Pi computer, which is then connected to the cloud.	This system is fail-safe and can be activated by downloading the average density in that area for a specific time period from the cloud.
2019	[7]	The input frame is a video sequence, and the Haar Cascade Classifier is used to recognise objects. A region of interest is chosen. The object is tracked by tracing the perimeter of the observed vehicle. Every passing vehicle within the ROI (Region of Interest) is tracked based on its position, and each new position is recorded as a new object to be tallied.	Every frame is compared to the previous frame; if the car appears in both frames and the difference in their coordinates is smaller than the maximum pixels,it is considered to be the same vehicle. If the difference is greater than the maximum pixels, we treat them as two independent automobile backgrounds.
2016	[8]	In order to achieve real-time performance they have combined the kalman filter with the Hungarian method.	To this purpose, detection quality has been discovered as a crucial influencing element in tracking performance, with altering the detector improving tracking by up to 18.9%This solution provides tracking precision in the same level as state-of-the-art online trackers while just using a simple combination of pre-existing techniques like the Kalman Filter and Hungarian algorithm for the tracking components.
2017	[9]	Directly add on top of the Simple Online and Realtime Tracking methods by adding a Kalman filter with a constant velocity motion	By conducting experiments it was concluded that addition of the features reduced the number of ID switches by 45%, achieving overall competitive performance at high frame

			rates.They also managed to achieve State-of-the-art performance while still being very fast.
2007	[10]	A two-level technique is offered. The method's lowest level uses an exponential forgetting algorithm to do background learning. The higher level analyses each pixel's Red, Green, and Blue (RGB) sequences and conducts dynamic pixel classification. The upper level determines the parameters of the background learning operation based on the pixel's class.	It has two improvements over previous methods. The backdrop pixel characteristics criterion can greatly reduce the fault judgments induced by minor video camera movements.The geometric properties of the lane line and road surface further strengthen the method's robustness.
2016	[11]	CNN is used for classification and localization s.YOLO is fast at processing images and detecting objects with its new approach.	A revolutionary approach to object detection at the time by Classifiers having been repurposed to do detection in previous work on object detection.
2012	[12]	An online learning approach, which develops the multi-target tracking problem as inference in a CRF model. Detection responses further are associated into tracklets.	An efficient algorithm is introduced to find associations with low energy, and the results of the experiment show significant improvement compared with current methods.
2020	[13]	Development of an adaptive model that combines YOLOv4 and Deepsort.	Results of the experiment show high accuracy and effectiveness for the proposed model.
2019	[14]	To generate a sequence of convolution feature images, the input image is fed into a pre-trained convolution layer. The closest feature photos to the original image are then chosen and blended into a new convolution feature image.	By modeling the main traits, the background subtraction method results in more accurate foreground item detection. Experiments revealed that the proposed solution for dynamic scenarios dramatically improved on standard background subtraction methods.

2009	[15]	An online technique for tracking multiple objects that is implemented in two stages. The particle filter with observer selection is used to construct a collection of reliable tracklets in the primary stage, which deals with a segment of the objects' occlusions. The detection feedbacks are taken from a transitory sliding window in the global stage to cope with uncertainty induced by full object occlusion and to produce a collection of candidate tracklets.	The results show that proposed implementation fundamentally enhances tracking veracity and efficiency in heavy observable surveillance environments with various occlusions.

YOLO Methodology



The figure above describes the basic workflow of YOLO, it breaks videos down to individual images and then feeds those images in. Taking a sample image in it divides the image to s * s grids. Create bounding boxes around all objects with confidences and simultaneously class probability Map. Finally, a non max suppression is applied and the final bounding boxes are obtained.

Conclusion

In this literature review, various object detection techniques like Fast R-CNN, Faster R-CNN, YOLO and object tracking methods like DeepSORT have been surveyed. After thorough research, it is found that YOLO stands out the most by being faster, accurate and less errant when compared to the other methods. We also found YOLO v4 to be the most suitable version . A study on papers showcasing the applications of YOLO in traffic counting based on density of traffic, and those including image processing has also been carried out. Many new object recognition and tracking approaches have been created in recent years that can improve the performance of our existing algorithms; we will continue to work on this in the future.

References:

[1] Ross Girshick, "Fast R-CNN", In Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2015, pp. 1440-1448

[2] Yuhao Xu and Jiakui Wang, "A unified neural network for object detection, multiple object tracking and vehicle re-identification", arXiv preprint arXiv:1907.03465v1 [cs.CV] 8 Jul 2019

[3] Anima Pramanik, Sankar K. Pal, J. Maiti , & Pabitra Mitra,"Granulated RCNN and Multi-Class Deep SORT for Multi-Object Detection and Tracking", IEEE Transactions on Emerging Topics in Computational Intelligence · January 2021

[4] Tuan Linh Dang, Gia Tuyen Nguyen & Thang Cao, "Object Tracking Using Improved Deep SORT YOLOV3 Architecture", ICIC Express Letters · October 2020

[5] Aderonke A. Oni et. al Nicholas Kajoh, "Video Based Vehicle Counting System for Urban Roads in Nigeria using YOLO and DCF-CSR Algorithms",International Journal of Engineering Research and Technology. ISSN 0974-3154, Volume 12, Number 12 (2019), pp. 2550-2558

[6] H. J. Naga, N. Nair, S. M. Jacob and J. J. Paul, "Density Based Smart Traffic System with Real Time Data Analysis Using IoT," 2018 International Conference on Current Trends towards Converging Technologies (ICCTCT), 2018, pp. 1-6, doi: 10.1109/ICCTCT.2018.8551108.

[7] N. Chhadikar, P. Bhamare, K. Patil and S. Kumari, "Image processing based Tracking and Counting Vehicles," 2019 3rd International conference on Electronics, Communication and Aerospace Technology (ICECA), 2019, pp. 335-339, doi: 10.1109/ICECA.2019.8822070.

[8]Alex Bewley, Zongyuan Ge, Lionel Ott, Fabio Ramos, Ben Upcroft "Simple Online and Realtime Tracking" 2016 IEEE International Conference on Image Processing (ICIP 2016)

[9]Nicolai Wojke, Alex Bewley, Dietrich Paulus "Simple Online and Realtime Tracking with a Deep Association Metric" 2017 IEEE International Conference on Image Processing (ICIP)

[10] Xiao-jun Tan, Jun Li, and Chunlu Liu, "A video-based real-time vehicle detection method by classified back-ground learning," World Transactions on Engineering and Technology Education 2007 UICEE Vol.6, No.1,2007

[11]Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi "You Only Look Once: Unified, Real-Time Object Detection"2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)

[12]B. Yang and R. Nevatia, "Multi-target tracking by online learning of non-linear motion patterns and robust appearance models," in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2012, pp. 1918–1925.

[13] Thanh-Nghi Doan and Minh-Tuyen Truong, "Real-time vehicle detection and counting based on YOLO and DeepSORT" 2020 IEEE 2020 12th International Conference on Knowledge and Systems Engineering (KSE)

[14]Tianming Yu , Jianhua Yang and Wei L "Refinement of Background-Subtraction Methods Based on Convolutional Neural Network Features for Dynamic Background" 2019 MDPI

[15] Junliang Xing; Haizhou Ai; Shihong Lao, "Multi-object tracking through occlusions by local tracklets filtering and global tracklets association with detection responses" 2009, IEEE