International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878, Volume-8 Issue-3, September 2019

Trajectory Object Detection using Deep Learning Algorithms

S Anitha Elavarasi, J Jayanthi, N Basker

Abstract*: Video surveillance data in smart cities needs to analyze a large amount of video footage in order to locate the people who are violating the traffic rules. The fact is that it is very easy for the human being to recognize different objects in images and videos. For a computer program this is quite a difficult task. Hence there is a need for visual big data analytics which involves processing and analyzing large scale visual data such as images or videos. One major application of trajectory object detection is the Intelligent Transport Systems (ITS). Vehicle type detection, tracking and classification play an important role in ITS. In order to analyze huge amount of video footage deep learning algorithms have been deployed. The main phase of vehicle type detection includes annotating the data, training the model and validating the model. The problems and challenges in identifying or detecting type of vehicle are due to weather, shadows, blurring effect, light condition and quality of the data. In this paper deep learning algorithms such as Faster R CNN and Mask R CNN and Frameworks like YOLO were used for the object detection. Dataset (different types of vehicle pictures in video format) were collected both from in-house premises as well as from the Internet to detect and recognize the type of vehicles which are common in traffic systems. The experimental results show that among the three approaches used the Mask R CNN algorithm is found to be more efficient and accurate in vehicle type detection.

***Keywords* : Deep learning, Intelligent Transport Systems, Mask RCNN, YOLO**

I. INTRODUCTION

Image processing is the process of analyzing and manipulating an image by using a computer algorithms in-order to extract useful information or to enhance the quality of the image. This can be used in the process of finding instances of the real world objects such as human, animals, birds, faces, bicycles, car, truck and building etc. Video content analysis (VCA) paves way to analyze and detect the various temporal and spatial details available in the video. Instead of analysis the still image, video contents can be analyzed in-order to extract useful insight from it. Insights from VCA can be used in various domains such as health care, home automation and automotive industry in-order to provide safety and security.

Trajectory Object detection and classification is a part of VCA. There is an urgent need for intelligent transport systems to replace human operation to monitor the surveillance area and to analyze a large amount of video footage in order to locate the people who are violating the traffic rules. Huge volume of video footage has to be analyzed in traffic control system especially in smart city application. There is a need for deep learning algorithm to easy the task. Deep learning is a part of machine learning methods based on artificial neural networks. Deep Learning algorithm such as Convolutional Neural Network (CNN)[9], Faster R-CNN [6] has been widely used in the field of object detection, segmentation and classification. The CNN approach consists of neurons, activation function and an output. Neuron contains learning units. The neurons in CNN receive several inputs, take a weighted sum over them, pass it through an activation function and finally respond with an output. The most common challenges associated with detection of vehicles are due to weather condition, shadows, blurring effect, light condition, type of vehicle and quality of the input data etc.

In this paper section 2 describes the related work and section 3 explains the architecture of Trajectory Object detection and classification system. Section 4 describes the results and discussion and finally section 5 concludes the work.

II. RELATED WORK

Yong Tang et al present vehicle detection and recognition. The author uses automatic monitoring digital cameras to take snapshots of moving motion pictures [1]. From the collected images, sequences are extracted to represent features of a vehicle. Initially machine learning algorithms like Haar-like feature and Adaboost algorithm are applied for feature extracting and constructing classifiers which is used to locate the vehicle over the input image. The drawback of the system is that it only works for day light image.

Viktoria Plemakova et al describe vehicle detection based on convolutional neural networks [3]. The main aim of the author is to train, classify and detect the vehicles from different angles using neural networks. The proposed CNN contains 6 convolutional layers and 5 max pooling layers. The challenges faced by the author are due to weather, light conditions and vehicle type diversity.

Yilmaz describes the vehicle detection using deep learning methods.

Manuscript published on 30 September 2019
* Correspondence Author
S Anitha Elavarasi*, Department of Computer Science & Engineering, Sona College of Technology, Salem, India. Email: anithaelavarasi@sonatech.ac.in
J Jayanthi, Department of Computer Science & Engineering, Sona College of Technology, Salem, India . Email: jayanthij@sonatech.ac.in
N Basker, Department of Computer Science & Engineering, Sona College of Technology, Salem, India . Email: baskern@sonatech.ac.in

The method takes five stages such as (1) loading the data, (2) the design of the Convolutional Neural Network, (3) training and configuration, (4) training of the R-CNN approach and (5) evaluation of the detector [4]. The parameters to be considered are image windows, selective search for object recognition, category independent object proposals, object segmentation using constrained parametric min-cuts and multiscale combinatorial grouping.
The approach achieves higher accuracy when the amount of training data is not restricted.

Yinghua li et al describes vehicle type detection based on compressed sensing and deep learning. The main aim is to extract the different features of vehicles from videos or pictures captured by traffic surveillance [2]. In-order to identify the type and classification of vehicles, vehicle saliency map and the convolutional neural-network (CNN) technique were applied. Compressed-sensing (CS) theory is applied to generate the saliency map to label the vehicles in an image. The concept of saliency map is used to search the image for target vehicles based on the use of minimization of redundant areas. The system is used for traffic monitoring and controlling purpose.

Dinesh Singh et al describe Traffic Monitoring in Smart City using visual Big Data Analytics[5]. Pre-processing process takes place as background subtraction method is applied on gray scale frames. Next step is to separate the objects in motion like bike, human, cars from static objects such as trees, roads and buildings. Object detection is done using Histogram of oriented gradients (HOG). Feature extraction is done using Scale invariant feature transform (SIFT). Map reduce function is used to schedule tasks and monitors the tasks.

III. PROPOSED APPROACH

Object detection is a process of localizing a particular segment. In-other words the process of identifying objects in an image and drawing bounding boxes around the object of interest to locate it within the image. There could be many bounding boxes representing different objects of interest within the image. Convolutional neural network is used to perform vehicle detection and classification. A naive approach to detect the various objects on an image is to take different regions of interest from the image, and use a CNN to classify the presence of the object within that region. The problem with this approach is that the objects of interest might have different spatial locations within the image and different aspect ratios. Therefore, algorithms like R-CNN, fast R-CNN (Fast Region-based Convolutional Network method), YOLO (You Only Look Once) have been deployed for detecting various object.
The General steps to be followed in object detection approach are:

- Data collection
- Annotating the data
- Training the model
- Validate the model
- Make prediction by running the model.

The proposed Trajectory Object Detection using Deep Learning approach (TODUDL) is shown in the figure 1. It uses the video data collected from the transport surveillance cameras. Pre- processing is performed by cleaning the dataset by using background subtraction method. Preprocessed data is given to various object detection algorithms (YOLO, Faster RCNN, Mask RCNN) for training and testing phase. These algorithms produce the segmented data as an output.
The steps involved in TODUDL are:

- Collecting the dataset
- In pre-processing background subtraction for removing the noisy data
- Implementation of object detection using YOLO framework.
- Implementation of Faster RCNN algorithm.
- Implementation of Mask RCN algorithm.
- Comparison of three object detection approach

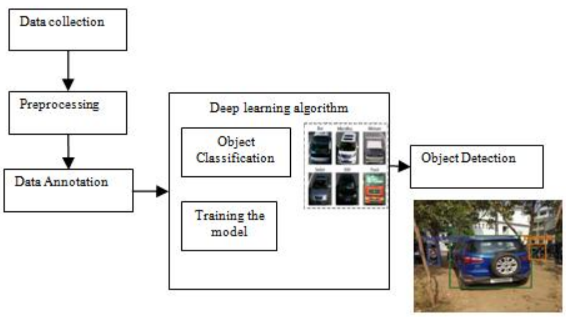

Fig 1. System Architecture of TODUDL

A. Data collection

The data set were collected both from real world and web repository.

B. Preprocessing

Background subtraction and labeling is performed in preprocessing phase. Background subtraction is carried out for removing noise and separating out foreground objects from the background in a sequence of video frames. This approach is used for detecting dynamically moving objects from static image. This is carried out by finding out the difference between the current frame and a reference frame. Labeling is performed using LabelImg. It is a graphical image annotation tool which is written in python, where the images will be manually separated and converted into .xml file for entering into the future applications. Labeled images include image name, height, width, class name, X-max, Y-max, X-min, Y-min values which is shown in the Table I.

Table I Labelling of Image

Image Name	Height	Width	Class name	Xmax	Ymax	Xmin	Ymin
Img1.jpg	640	480	Bus	378	75	521	226
Img2.jpg	640	480	Car	88	71	232	224
Img3.jpg	640	480	bus	209	96	354	178

C. Faster RCNN Algorithm

An object detection algorithm that eliminates the selective search algorithm and lets the network learn the region proposals like similar to Fast R-CNN algorithm, the image will be provided as an input to a convolutional network which provides a convolutional feature map. Instead of using

Retrieval Number: C6564098319/2019©BEIESP
DOI:10.35940/ijrte.C6564.098319
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

selective search algorithm on the feature map to identify the region proposals, a separate network is used to predict the region proposals [6]. The predicted region proposals is then reshaped using a Region of Interest (RoI) pooling layer which is then used to classify the image within the proposed region and predict the offset values for the bounding boxes.

Steps to be followed for faster RCNN algorithm

- Read the input data
- Extract the required region by generating a bounding box
- Compute the CNN features
- Classify the various regions.

D. Mask RCNN algorithm

The MASK RCNN [7], [8] is just the combination of a Faster R-CNN and FCN (Fully Convoluted Network) in one mega architecture that performs pixel wise boundary that does object detection where object detection is the combination of class and bounding box. Steps to be followed for mask RCNN algorithm:

- Read the input data
- Extract the required region
    - Creation of bounding box
    - Class identification
- Pixel wise boundary identification using binary mask
- Classify the various regions

Fig 2.Object detection using mask RCNN

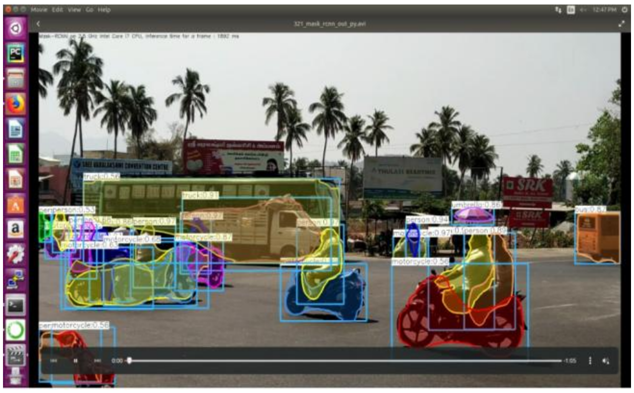

Fig 3. Object detection using Mask RCNN for video input

E. You Only Look Once

YOLO (You Only Look Once) is one of recent object detection framework which is extremely fast and accurate[10]. YOLO uses the whole image in a single instance instead of separate regions to make decisions predicts the bounding box. The network architecture is based on GoogleLeNet where inception modules are replaced by reduction layers. The given image is divided into an S x S grid. Bounding boxes are drawn around images. The bounding boxes are weighted by the associated probabilities. Objects are classified based on classifiers chosen by the user. Twenty convolutional layers are first pre-trained and then another four convolutional and two fully-connected layers are added to the model. Steps to be followed in YOLO are:

- Load the pre trained weights
- Read the input data resized to have the same width and height as the first layer of the neural network
- Assign Threshold value (Intersection Over Union (IOU) and Non-Maximal Suppression (NMS))
- Detect the objects and classify the various region and display their probabilities

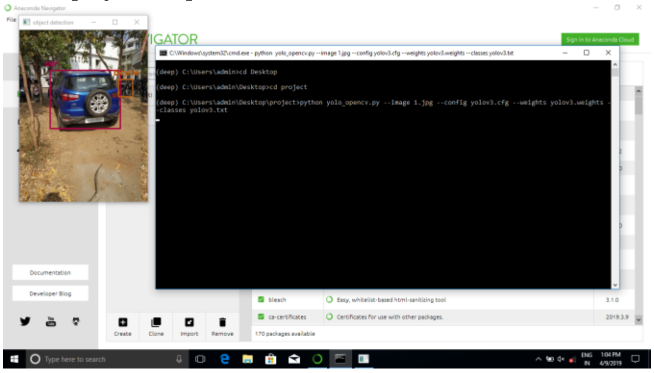

Fig 4. Object detection using YOLO

IV. RESULT AND DISCUSSION

A. Dataset

The data set were collected both from real world and web repository. Real world data from our college premises (capturing data from the surveillance cameras in the front gate). The web repository are (1) COCO (Common Object in COntext) and its URL is http://cocodataset.org and (2) ImageNet and its URL is http://www.image-net.org/. The data is been divided in the ratio of 60:40 for training and testing data respectively. The training data is used for training the various object detection algorithms and test data is used for evaluate the object detection algorithms.

B. Experimental setup

The experiments are conducted on a cluster of two machines running Ubuntu 16.04 having specifications Intel Core i5-4590 Processor 3.3Ghz 6M Cache, 16 GB Memory, 500 GB SATA HDD, 10 Gbe Ethernet Card, 18.5” Monitor x CPU Processor. It uses Jupyter Notebook, Keras Framework, Numpy, Seaborn, Pill, Pillow, MatPlot, h5py, Opencv, Pandas, Tensorflow, StatsModel, Scipy and Scikit- learn.

C. Result comparison

The main goal of this trajectory object detection is to extract the different features of vehicles from videos captured by traffic surveillance and to identify the types of vehicles using deep learning algorithm. Figure 5 shows the comparison of object detection using Faster R CNN, YOLO and Mask R CNN algorithm. Object classification is performed using the pre-trained model ImageNet [12]. ImageNet an image database consist of more than 1000 of images of various categories. The Object detection accuracy for different type of vehicle such as car, bus, motorcycle and truck were compared among the three deep learning approaches.

Faster RCNN achieves a minimum of 51% and a maximum of 70% and YOLO gives a minimum of 50 and a maximum of 62% accuracy. The result clearly shows that mask RCNN algorithm gives an accuracy of nearly 86.25% with a minimum of 82% to a maximum of 91% for different types of vehicle available on the test dataset when compared with other two algorithms.

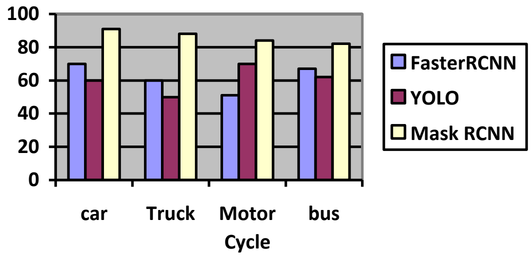

Fig.5. Comparison Faster R CNN, YOLO and Mask R CNN

V. CONCLUSION

The traditional computer vision techniques are unable to analyze huge amount of visual data generated in real-time. With the development of computer vision techniques and consequent accessibility of video image data, new applications have been enabled to on-road vehicle detection algorithms. The application works well in the field of (1) Traffic Trajectory Tracking, (2) Identification of unlicensed Vehicle and (3) Capturing Traffic Violators. The proposed Trajectory Object Detection using Deep Learning approach (TODUDL) approach uses the deep learning algorithms like Faster R CNN and Mask R CNN and frameworks like YOLO. The main aim of this paper is to successfully train the system to detect any vehicle from a video captured by the traffic surveillance camera using Faster RCNN, Mask RCNN and YOLO framework. After successful training, performance of all the three algorithms are compared on vehicle detection (i.e how accurately it detects the type of vehicle). The average accuracy for Faster RCNN, Mask RCNN and YOLO framework were compared. An accuracy of nearly 86% is achieved by mask RCNN algorithm for varied types of vehicle

ACKNOWLEDGMENT

The authors would like to thank the Department of Science and Technology (DST), India for their financial Assistance through Science for Equity Empowerment and Development (SEED) programme (Grant No: SEED/WS/018/2015). The experimental analysis performed in Data Analytics and Solutions lab (Catalysed & Supported by SEED division, DST, New Delhi,) Sona College of Technology, Salem, Tamil Nadu, India.

REFERENCES

1. Tang, Yong, Congzhe Zhang, Renshu Gu, Peng Li, and Bin Yang. "Vehicle detection and recognition for intelligent traffic surveillance system." Multimedia tools and applications 76, no. 4,,2017, pp: 5817-5832.
2. Li, Y., Song, B., Kang, X., Du, X., & Guizani, M, Vehicle-type detection based on compressed sensing and deep learning in vehicular networks. Sensors, 18(12), 2018, pp:4500.
3. Plemakova, V. Vehicle Detection Based on Convolutional Neural Networks.
4. Yilmaz, Abdullah Asim, Mehmet Serdar Guzel, Iman Askerbeyli, and Erkan Bostanci..,"A Vehicle Detection Approach using Deep Learning Methodologies." arXiv preprint arXiv:1804.00429, 2018.
5. Singh, Dinesh, C. Vishnu, and C. Krishna Mohan. "Visual big data analytics for traffic monitoring in smart city." 2016 15th IEEE international conference on machine learning and applications (ICMLA). IEEE, 2016, pp. 886-891.
6. Ren, S., He, K., Girshick, R., & Sun, J, “Faster r-cnn: Towards real-time object detection with region proposal networks”,In Advances in neural information processing systems , 2015, pp. 91-99.
7. He, K., Gkioxari, G., Dollár, P., & Girshick, R, “ Mask r-cnn”, In Proceedings of the IEEE international conference on computer vision, 2017, pp. 2961-2969.
8. https://www.analyticsvidhya.com/blog/2018/07/building-mask-r-cnn-model-detecting-damage-cars-python/
9. https://www.analyticsvidhya.com/blog/2018/12/guide-convolutional-neural-network-cnn/
10. https://www.analyticsvidhya.com/blog/2018/12/practical-guide-object-detection-yolo-framewor-python/
11. http://cocodataset.org.
12. http://www.image-net.org/

AUTHORS PROFILE

S Anitha Elavarasi is working as an assistant professor in the department of Computer Science and Engineering, Sona College of Technology (SCT), Salem, Tamil Nadu, India. She completed her BE CSE from Bharathiyar University, ME CSE from SCT, Anna University and PhD from Anna University. She is having 13 years of teaching experience. Her area of interest includes Data mining and Big Data Analytics, Artificial Intelligence, Internet of Things. She has published more than twenty papers in the national and international journals and conferences. She has received many appreciation awards for the teaching and learning and Research and Development activities.

Dr. J Jayanthi is working as an associate professor in the department of CSE, Sona College of Technology, Salem, Tamilnadu, India. She is having three years of Industrial experience and 17 years of teaching experience. Her area of interest includes web technologies, software engineering, AI, Data mining and Big Data Analytics. She has published more than fifty papers in the national and international journals and conferences. She is one of the co-investigator of the DST funded project on CDSS for Breast Cancer for rural women community. She has received many appreciation awards for teaching and learning and R&D activities.

N.Basker is working as an assistant professor in the department of Computer Science and Engineering, Sona College of Technology (SCT), Salem, Tamilnadu, India. He completed his BE CSE from Paavai Engineering College, Anna University, ME Software Engineering from Sona College of Technology, Anna University. He is having 10 years of teaching experience in SCT. His area of interest includes software engineering, Software Testing Data mining and Big Data Analytics, Internet of Things. He had published more than ten papers in the national and international journals and conferences.