Pedestrian Detection using Machine Learning and its Comparison with HOG and NMS

Rahul Subramaniam[1], Anmol Gaba[2], Shashank R B[3], Rashmi R[4], Dr. G N Srinivasan[5]

[1-3]Department of Information Science and Engineering, R.V College of Engineering, Bangalore, Karnataka, India
[4]Professor, Department of Information Science and Engineering, R.V College of Engineering, Bangalore, Karnataka, India

---

Abstract - *Pedestrian detection is an essential component of image and video surveillance system, on account of its provision of information with respect to the semantic aspects of understanding video footage. Its potential applications can be best stated in the field of automotive engineering, as well as improvement of safety systems. A plenitude of Car manufacturers offer this functionality as a part of their Advanced Driver Assistance System. In this work, the dataset, obtained from the MIT People's data, involves a set of images, that are then bifurcated into the obvious training set and testing set. Prior to the the division, appropriate image processing is performed in order to get the images to the requisite size and format. The training and testing set are then tested against the Machine Learning algorithms and the Histogram of Oriented Gradients (HOG) using Support Vector Machines. The results obtained are then compared with the responses of Pedestrian detection using Non Maximum Suppression Algorithm (NMS) in order to estimate how accurate the adopted pedestrian detection approach is.*

Key Words: Detection, Support Vector Machine, Maximum Suppression, HOG Descriptor, Machine Learning.

1. INTRODUCTION

Pedestrians on road are prone to accidents and in some cases may even be the cause of accidents. The National Highway Traffic Safety Administration under the United States Department of Transportation [1] states that the number of pedestrians killed in traffic clashes has increased by a near 3 percent in the year 2018. This led to a total of 6,283 deaths: The maximum number of deaths since the year 1990. There are several traffic awareness programs that have been planned by various governmental and non-governmental organizations organized in order to help mitigate the number of casualties resulting out of road accidents.

Leveraging the power of machine learning is a viable option to detect vehicles and pedestrians on road. It can help notify pedestrians or vehicles with respect to how close the pedestrian/vehicles are, how many of them are approaching etc. This is done by classifying the images we have into two classes: pedestrian positive and pedestrian negative.

Pedestrian negative images are images that do not have any pedestrians in them while pedestrian positive images are images that have 1 or more pedestrians in their contents.

Machine Learning algorithms are broadly classified into Supervised and Unsupervised Learning Algorithms. Supervised learning algorithms are those which involve learning of a function that maps the input to output, forming input output pairs. They are generally of two types: Classification(use of categorical class variables) and Regression (predict real numbered outputs). This project makes use of the Classification technique. Unsupervised learning algorithms on the other hand are used to draw inferences from input data without labeled responses. They are divided into Clustering and Association. This project employs a clustering algorithms to perform detection.

Models of this kind can help save as many lives as quickly as possible and turn out to be an efficient and cost-effective solution. It can be further improved by the inclusion of features such as pedestrian description. Such a system has important applications in Self driving vehicles too.

2. LITERATURE SURVEY

On account of the multiple advantages that pedestrian detection offers to the community, it has gained a lot of traction in recent times and a tremendous amount of research is being done in the field of Pedestrian detection. The number of approaches are manifold, each done with the aim of being more accurate than the previous approaches. These approaches primarily deal with the use of Machine Learning and Neural Networks.

Pedestrian detection was done using a newly proposed algorithm called the Regional Proposed Network.It entails the use of a Pedestrian Retrieval framework with R-CNN. The similarity between two feature vectors is measured accurately using a linear combination of the absolute difference and product of the elements of the two vectors [2].

Another perspective to Pedestrian Detection was done using a Single Shot Detector framework that deploys different activation maps using OpenCV to achieve a reliable pedestrian detection. It incorporates deep learning to achieve its purpose and is primarily designed for real time operations [3]. This is supplemented with the need of high accuracy, that the paper claims to be a daunting task.

A third approach to Pedestrian Detection dealt with the use of a CENTRIST feature extraction in combination with offline Support Vector Machine Classifier in order to perform the Detection[4]. It compares the CENTRIST AND HOG approaches to feature extraction and eventually claims CENTRIST to be the more accurate of the two approaches. It emphasizes on the use of edge classification to deepen contour feature.Elimination of background noise and local texture are also taken care of in this work. Finally, it offers comparison of the detection performance based on the INRIA Pedestrian Detection Dataset.

A Monocular Vision approach was used to perform Motion Pedestrian Detection. The approach analyzes pedestrian movement. The algorithm design and implementation process of detecting and tracking pedestrian movement is studied in detail. Post validation, the approach is deemed to be effective[14].

A relatively older approach dealt with the application of Neuromorphic Visual Processing in the field of Pedestrian detection. The approach mimics visual systems to produce relevant results. It is based on Hubel and Wiesel's experimentation and Hodgkin Huxley Formalism [15].

Pedestrian Detection was also performed using Radiometric Temperature Information [5]. Conventionally, the brightness of thermal images is used for Pedestrian Detection. This work deems the conventional approach unstable in its nature, leading to background distortion. This paper shows preference towards the use of Calibrated Temperature over brightness as a parameter. An ACF Boosted tree detection platform is used to validate the effectiveness of the Radiometric Temperature Approach.

A final approach this paper mentions, dealt with the use of ultrasonic signals to perform Pedestrian Detection[6]. It deals with distinguishing a pedestrian detected by the driver assistance system from a car detected by the system. It incorporates pedestrian protection systems in the case of a collision. Sensors present evaluate the electronic signals that are back scattered from the obstacle. Components of such a system involve a transducer, a component for transducer control, signal acquisition system, and a computer that performs signal analysis.

3. DATA DESCRIPTION

As shown in the below diagram, the process of obtaining and defining the data is mainly composed of three essential steps: Data Sourcing, Positive data sample creation, Negative data sample creation.

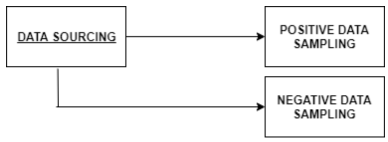

Fig. 1. Flow diagram of the data description step

A. Data Source

The dataset used in the given project is the MIT People's dataset. The dataset is open sourced in its nature and is obtained from [7]. With a tar.gz format compression, the dataset is 10 MegaBytes Compressed and 22 MegaBytes Uncompressed.

The dataset has a total of 924 colour images of pedestrians. Each of the images is scaled to a 64*128 pixels and aligned so that the pedestrians body lies in the centre of the image. Aligning and Scaling the data samples is a step of high level of importance for this given dataset. The dataset was split into eighty percent training data and twenty percent testing data. The above ratio (4:1) was used since the size of the dataset is small (a mere 924 images) and hence necessitates the use of greater percentage of images to train the data while the random state value chosen is forty two.

B. Positive Data Sample Creation

This step primarily deals with loading the set of 500 pedestrian positive images (i.e Images that have zero or more Pedestrians present in it) into the NumPy array. Before having to put the images into the array, we use the OpenCV Library functions (available as a part of Python) to reduce the given image from 64*128 pixels to 64 * 64 images, so that to remove any background and incorporate only the pedestrians.

C. Negative Data Sampling

The remaining images from the dataset are pedestrian negative images, i.e. images without any pedestrian and with just plain pathways. These images were first scaled from their original size to a 64*64 size using the OpenCV Library of Python, followed by their addition into another numpy array.

Images with no pedestrians were marked with a zero in the Y_NEG(array indicating pedestrian present or not) numpy array, while images with pedestrians were marked with the value indicating the number of pedestrians in the Y_POS (array indicating number of pedestrians, a positive number). These two arrays, along with the X_POS and X_NEG arrays (that store the individual images) were coalesced to obtain a singular X(image array) and Y(value to be predicted array) array. This in turn was split into the training and testing set in accordance with the percentages written in the data sourcing step of this paper.

International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 08 | Aug 2020 www.irjet.net p-ISSN: 2395-0072

4. METHODOLOGY

This covers the technique and flow of events that were used to perform the detection process. The prediction methodology itself is composed of two integral steps: applying the descriptor/algorithm to the data model, the final prediction process.

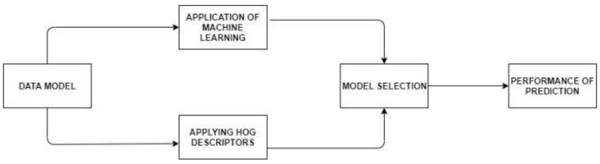

Fig. 2. Flow diagram of the methodology step

A. Model using Machine Learning

For any given prediction/detection problem, there are numerous Machine Learning Algorithms that can be used. Thus, to find out the one most suitable to our purpose, all of them must be evaluated on the same data based on a suitable parameter.

For this given Data Model, we chose the accuracy_score metric that is provided by the sklearn library of python as our evaluation metric[10]. The accuracy score is calculated as follows

(1)

In the above formula, y refers to the set of predicted values for a given test set, while y' refers to the actual values of the test set that are being predicted. The number of samples in this case is n_samples. Both y and y' contain the values -1,1, that indicate if there is a pedestrian present in this given image or not.

However, an anomaly observed in case of object detection in images using Machine Learning Algorithms was that the X_Train and X_Test (The Training and testing set with all parameters) is three dimensional in nature. This was in contrast to the fact that most Supervised Learning Algorithms tend to operate on two dimensional X_Train and X_test. Hence, the X_train and X_test matrices need to be reshaped that results in an incorrect technique to detect the pedestrian. For example, the Logistic Regression algorithm delivered an accuracy of 100% when operated on a two dimensional X_Train. However, this was proved to be wrong when tested against a single random image. The same was observed in case of K Neighbors Classifiers.

TABLE I. ERRONEOUS ACCURACY OBSERVED BY SUPERVISED MACHINE LEARNING ALGORITHMS

Algorithm	Score
Logistic Regression	1.0
Decision Tree Classifier	0.9178
K-Nearest Neighbors (n=2)	1.0
K-Nearest Neighbors (n=8)	0.8823

The standard three dimensional X_train and X_test worked with the Support Vector Machine (provided as a part of the OpenCV library in Python), and also with a combination of K-Means Clustering and Random Forest Classifier. The observation of using K-Means and random forest classifier was affirmed by [8]. The accuracy attained were as follows

TABLE II. ACCURACIES OF MODELS WORKING ON THE 3 DIMENSIONAL X-TRAIN

Algorithm	Score
Support Vector Machine(OpenCV)	0.7956
K-Means Clustering and Random Forest	0.8328
Histogram of Oriented Gradients	0.981

B. Applying Histogram of Oriented Gradients Descriptor

With a given image from the set I, there is a sliding window that generates a set of patches. The Histogram of Oriented Gradients Descriptor is calculated for each of these patches, resulting in a series of feature vectors describing the contents of this patch. It is available as a part of the OpenCV library in Python[11]. The overall HOG Descriptor is calculated using the image gradient orientation of each pixel.

The parameters provided to the HOG Descriptor were: window size (48,96), block size(16,16), block stride (8,8),cell size(8,8), number of bins (9). The histogram orientations are stored in these 9 bins for each image cell. The blocks are created in order to avoid variance due to illumination.

In this process, we created a HOG Descriptor object and set an SVM Detector to it using functions available as a part of the OpenCV library. The HOG Descriptor has its own People Detector function, that identifies persons on the basis of certain features. This People Detector function is applied to the SVM Detector function. The HOG Descriptor approach managed to attain an overall accuracy of 98 percent (i.e. it was able to rightly detect whether a pedestrian was there in the image or not 98% of the time).

C. Prediction

From the Application of Machine Learning side, Support Vector Machine and K-Means Clustering approaches were selected, while from the Histogram with Oriented Descriptors, the Descriptor with SVM Detector was chosen. Overall, since the HOG Descriptor yielded a greater accuracy, it was the approach adopted for prediction

5. RESULTS AND DISCUSSIONS

Non Maximum Suppression technique is a well established object detection library, and is useful for Pedestrian Detection in crowded scenes[9]. The Non Maximum Suppression technique uses a list of proposal boxes, and filters them to obtain a new proposal set based on corresponding confidence scores and a predefined threshold. It uses an Intersection to Union ratio to measure the overlap between two proposals[13].

While implementing Non Maximum Suppression on the pedestrian detection model, additional penalties were imposed in order to produce compact boxes, so that the model is less sensitive to the Non Maximum Suppression threshold. and hence comparing the results obtained by the project with that of the Non Maximum Suppression would give an idea with respect to how correct and accurate the detection is. This given work presents a comparison of the results obtained by HOG Descriptors (as concluded in the previous section) and Non Maximum Suppression technique for two given images, as shown in Fig 3, Fig 4 and Fig 5,Fig 6. The basis of the comparison used is the number of bounding boxes that contain pedestrians in them. Greater the number of bounding boxes(with pedestrians in them), greater is the precision of the approach, and hence more suitable is the approach for real world applications.

Fig. 3. Result of Non Maximum Suppression for the first test image

Fig. 4. Result of Histogram Of Oriented Descriptors Graph for the first test image

In Fig 3 it was observed that the Non Maximum Suppression technique managed to create five bounding boxes. This was in contrast to the Histogram Oriented Descriptor that managed to create only two bounding boxes, as shown in Fig 4. Thus for the first test image, Google Vision Object detection is more precise in nature.

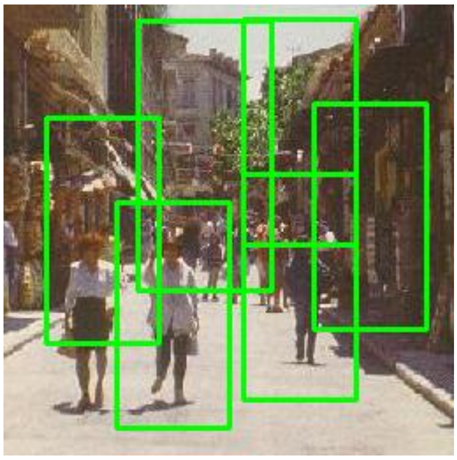

Fig. 5. Result of Non Maximum Suppression for the second test image

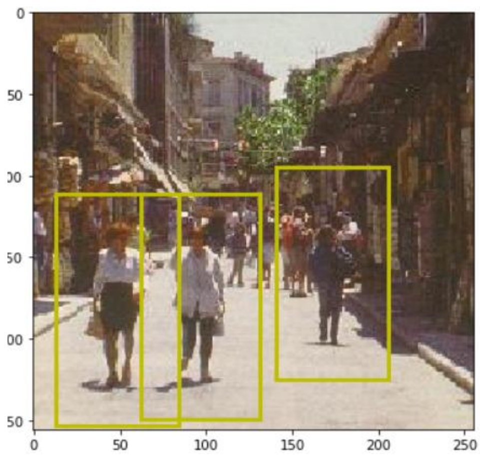

Fig. 6. Result of Histogram Of Oriented Descriptors Graph for the second test image

In Fig 5, it was observed that the Non Maximum Suppression technique managed create five pedestrian bounding boxes, while the Histogram of Oriented Gradient Descriptors yielded only three bounded boxes. Thus in this case also, it is safe to assume that Google Vision Object detection was more precise than the Histogram of Oriented descriptors approach in this case also.

5. CONCLUSION

This work presents that among the Machine Learning approaches, an Unsupervised Learning approach of K-Means Clustering along Random Forest Classifier algorithm has the highest accuracy, while the Support Vector Machine algorithm has been observed to have the second highest accuracy in detecting pedestrians. Support Vector Machine also is the most accurate Supervised Machine Learning algorithm in detecting pedestrians. The accuracy was calculated using the sklearn provided accuracy_score function[10]. However, none of the Machine Learning algorithms used are as accurate as the approach Histogram of Oriented Gradients object along with a Support Vector Machine detector. Thus, despite trying and testing of multiple algorithms, the HOG Descriptor approach was adopted. Finally, in order to check how precise was the HOG Descriptor approach when trying to detect the number of pedestrians, it was found that the Non Maximum Suppression technique was able to detect almost all the pedestrians. Probable causes for the lack of precision is the fact that Non Maximum Suppression imposes penalties to ensure more compact bounding. Applications of this work are primarily in the field of Safety systems, Traffic violations and Driver Assistance systems in vehicles.

REFERENCES

[1] United States Department of Transportation, National Highway Traffic Administration "https://www.nhtsa.gov/road-safety/pedestrian-safety"

[2] Enjia Chen, Xianghong Tang and Bowen Fu "A Modified Pedestrian Retrieval Method based on faster R-CNN with Integration of Pedestrian Detection and Re-Identification" 2018 International Conference on Audio Language and Image Processing,

[3] Zahid Ahmed; R. Iniyavan and Madan Mohan P, "Enhanced Vulnerable Pedestrian detection using Deep Learning," in 2019 International Conference on Communication and Signal Processing .

[4] Zhijie Huang, "Pedestrian Detection Algorithm in Video Analysis based on Centris" 2016 International Conference on Intelligent Transportation, Big Data and Smart Cities.

[5] Sungho Kim, Taehwan Kim "Radiometric Temperature-based Pedestrian Detection for 24 Hour Surveillance" 2018 IEEE International Conference on Multimedia and Expo Workshops.

[6] Andreas H. Pech,Peter M. Nauth, Robert Michalik "A New Approach for Pedestrian Detection in Vehicles by Ultrasonic Signal Analysis" IEEE EUROCON-2019, 18th International Conference on Smart Technologies

[7] MIT People's Dataset for non commercial purposes.The Hackerearth platform official website "http://cbcl.mit.edu/software-datasets/PedestrianData.html"

[8] ResearchGate website k-means clustering pedestrian detection"https://www.researchgate.net/publication/280001413_Pedestrian_classification_using_K-means_and_Random_Decision_Forests/link/57161f9c08ae6177286d3b97/download"

[9] Website blog "https://towardsdatascience.com/pedestrian-detection-using-non-maximum-suppression-b55b89cefc6"

[10] Sci-kit documentation on calculating Accuracy score "https://scikit-learn.org/stable/modules/generated/sklearn.metrics.accuracy_score.html"

[11] OpenCV documentation on Histogram Oriented Gradient Descriptors "https://docs.opencv.org/3.4/d5/d33/structcv_1_1HOGDescriptor.html"

[12] OpenCV official website with mention of SVM Detector in HOG Descriptor"https://docs.opencv.org/2.4/modules/gpu/doc/object_detection.html"

[13] Website blog "https://towardsdatascience.com/non-maximum-suppression-nms-93ce178e177c"

[14] Yi Lei; Zhijie Huang "Research On Pedestrian Detection using Monocular Visiion" in 2018 International Conference on Robotics and Intelligent Systems.

[15] Song Han, Woo-Sup Han "Application of neuromorphic visual processing in pedestrian detection technology" in 2013 Science and Information Conference.