MIDSEM REPORT

Abstract

Self-driving vehicles are a promising technology that can improve transportation safety, efficiency, and convenience. However, they pose significant challenges regarding object detection and intelligent decision-making in complex and dynamic road environments. This project aims to develop a state-of-the-art system that can recognize objects, animals, and pedestrians on the road and make informed decisions to navigate safely and effectively in various scenarios, such as an incoming vehicle with a specific speed or a flying ball approaching the vehicle. The project focuses on training the model for Indian traffic and infrastructural problems, often different from those in other countries. The project uses a combination of deep learning models, such as YOLO, Faster R-CNN, and HOG + SVM, to achieve high accuracy, speed, and robustness in pedestrian detection and intent forecasting. This report sheds light on the present progress of the project.

Introduction

Due to the formidable nature of the roads in many poorer sections of the world, existing models in self-driving cars are not suitable for every scenario, as they may not be able to handle various variations like pose, scale, illumination, and background. There is an immense need to improve the current models to incorporate more variables and to allow them to make better predictions.

Self-driving vehicles are an emerging technology that can enhance transportation safety, efficiency, and convenience. However, they face significant challenges in object detection and intelligent decision-making in complex and dynamic road environments. This project aims to develop a cutting-edge system that can identify objects, animals, and pedestrians on the road and make intelligent decisions to navigate safely and effectively in various scenarios, such as an incoming vehicle with a specific speed or a flying ball closing the vehicle. The project concentrates on training the model for Indian traffic and infrastructural problems, often different from those in other countries.

The project focuses on enhancing holistic object detection and informed decision-making for self-driving vehicles, as the current benchmarks are not calibrated for vast Indian scenarios and are susceptible to errors and misjudgments.

Literature Survey

1) *Pedestrian Crossing Intention Forecasting at Unsignalized Intersections Using Naturalistic Trajectories*

- Paper titled “Pedestrian Crossing Intention Forecasting at Unsignalized Intersections Using Naturalistic Trajectories” by Esteban Moreno, Patrick Denny, Enda Ward, Jonathan Horgan, Ciaran Eising, Edward Jones, Martin Glavin, Ashkan Parsi, Darragh Mullins, and Brian Deegan
- The article emphasizes the reliance of current autonomous vehicle designs on onboard sensors, which may have a limited field of view, leading to potentially dangerous situations.
- Uses dataset of “Dataset of Naturalistic Road User Trajectories at German Intersections.”
- Pre-processing steps- 1) labeling pedestrian intention 2) imbalance handling 3) feature selection 4) training and testing split 5) handling the unclear scenario
- Models used - random forest, feed-forward neural network
- The Random Forest model is the primary focus and outperforms the neural network in terms of accuracy and training times in the context of the described study.

2) Histograms of Oriented Gradients for Human Detection

- Paper titled “Histograms of Oriented Gradients for Human Detection” by Navneet Dalal and Bill Triggs
- The method is based on evaluating well-normalized local histograms of image gradient orientations in a dense grid.
- Uses a dataset from the MIT pedestrian database containing 509 training and 200 test images of pedestrians in city scenes.
- Pre-Processing Steps- 1) color space conversion 2) gradient computation 3) normalization 4) region of interest (ROI) selection 5) noise reduction
- Models used - Linear SVM, HOG, SIFT
- HOG was a better model as SIFT features are not well adapted to human detection problems. Also, HOG has a higher detection rate and lower false positives.

3) Pedestrians’ Detection Methods in Video Images

- “Pedestrians’ Detection Methods in Video Images: A Literature Review”
- This paper reviews current algorithms for pedestrian detection using image processing, where used images have been obtained from video surveillance or conventional cameras
- Traditional Methods Traditional methods for pedestrian detection often involve handcrafted features and classifiers. Features such as Histogram of Oriented Gradients (HOG), Scale-Invariant Feature Transform (SIFT), and Speeded Up Robust Features (SURF) have been widely used.
- Models used - Linear SVM, HOG, SIFT

- Pre-processing Steps- 1) Image smoothing 2) Region of interest selection 3) Color Transformation 4) Inverse Perspective mapping 5) Segmentation
- These steps are common in many pedestrian detection methods and help to prepare the image data for further processing and analysis.

Dataset

We have reviewed several online available pedestrian detection datasets to fast-track the data collection process and increase its volume multifold. While several datasets were barred behind registrations and pending access, or some websites showed server error access, we were able to acquire data from the following sites to utilize them for our use case:

1. http://zheng-lab.cecs.anu.edu.au/Project/project_prw.html - 11,816 images

2. https://www.cis.upenn.edu/~jshi/ped_html/ - 170 images with pedestrian size falling into [180,390] pixels

3. http://www.cbsr.ia.ac.cn/users/sfzhang/WiderPerson/ - 13,382 images

While the above data was initially intended for pedestrian detection, we can utilize them for our purpose of traffic detection till we get access to a better dataset or gather our data. This would act as a stepping stone to detect and form models before applying the methods to a vast dataset. It would allow us to find the most suited model for our project.

Feature Selection and Output:

We have identified the individual objects and their types (pedestrians, bikers, and more) along with their initial and final coordinates and postures to indicate standing still or walking, moving, and running.

Pre-processing Steps

1) Removing Outliers
2) Importing necessary dataset
3) Checking for any missing/corrupted data
4) Converting images/videos in a suitable format
5) Feature selection
6) Dataset training and testing split
7) Normalization

Methodology

Models Used

- Histograms of Oriented Gradients

- Support Vector Machines
- Logistic regression
- Gaussian Naive Bayes
- Random Forest

Model Details

Logistic Regression:

- We used this model to predict if there is any object in the front of the vehicle or not.
- Furthermore, this model can predict whether the vehicle will move left or right.

Histograms of Oriented Gradients and SVM:

- It uses histograms of oriented gradients to capture the shape and appearance of an object in an image. It works by dividing the image into small cells, computing the gradient magnitude and direction for each pixel in the cell, creating a histogram of gradient orientations for each cell, normalizing the histograms using blocks of cells, and concatenating the histograms into a feature vector.
- We have used various normalization methods like L1-Norm and L2-Norm for the contrast and invariance of the HOG features.
- We have used SVM as the classifier for giving binary decisions as output

Random Forest:

- It uses decision trees to perform classification or regression t. It works by creating multiple decision trees from random subsets of the training data and combining their predictions by classification or regression. We used 100 trees. Using more trees can improve accuracy and reduce variance but increase the computation time and complexity.
- Used the default value “None” in the max depth of tree, meaning trees will grow until all leaves are pure or contain less than a minimum no. of samples.
- Used square root of total no. of features for classification and divided by three for regression.
- Used Gini impurity for classification and mean squared error for regression.

We have also measured the Accuracy, Precision, Recall, and F1 Score for each method used in the model. In addition, we have used various hyperparameters and tried to find some optimal parameters to improve the model's accuracy.

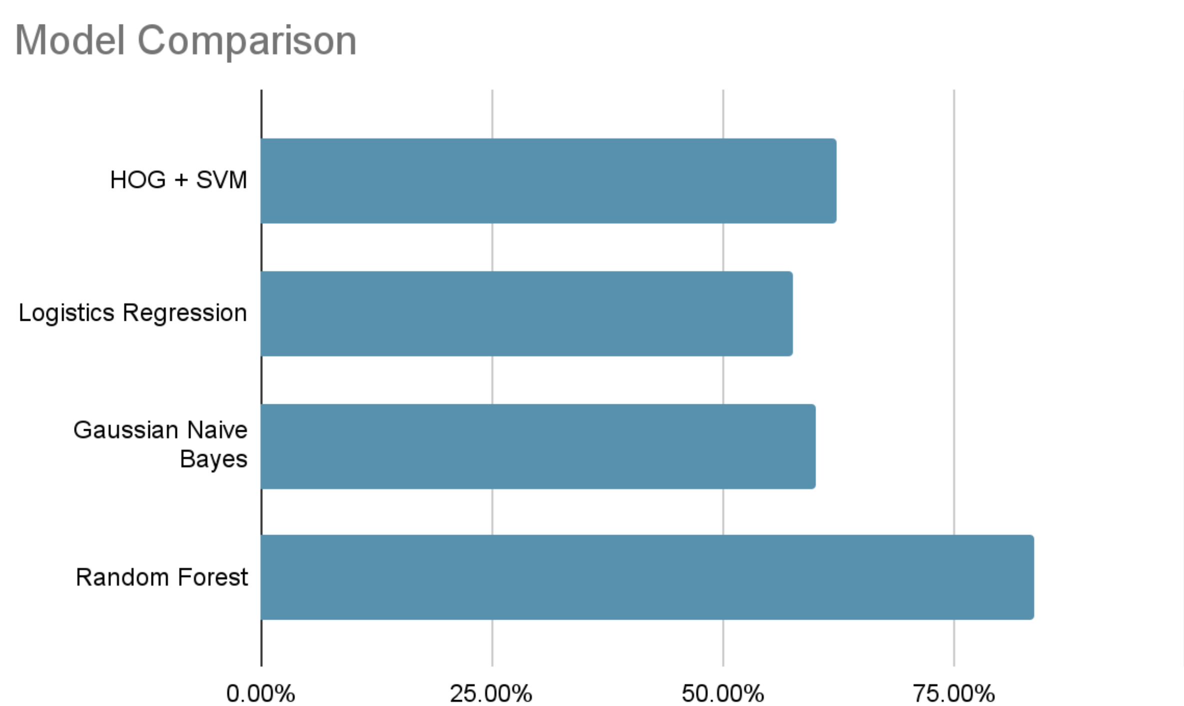
Results And Analysis

Findings

Model	Accuracy	Precision	Recall	F1 Score
HOG+SVM	62.38%	0.56	0.51	0.44
Logistic Regression	57.61%	0.50	0.45	0.42
Gaussian Naive Bayes	60.02%	0.55	0.52	0.53
Random Forest	83.61%	0.80	0.80	0.80

Analysis

1. HOG + SVM is a simple and fast method but less robust to variations.
2. Naive Bayes is a quick method for text but is not suitable for image data.
3. Random forest can perform multiclass classification and provide probability over prediction, but it could be more efficient for high-dimensional data.



Conclusion

Learnings

- The project helped us learn and use diverse machine-learning models with various hyperparameters.
- It helped us learn how to evaluate various models based on their precision, recall, F1 Score, and accuracy metrics.
- It allowed us to explore various models for the project and assisted with the next step of narrowing down our technique.

Future Tasks

- The future scope of the project can incorporate finding and experimenting with some unexplored models.
- Better training of the data further optimizes the hyperparameters and combines all the better models to get even better performance.
- Using methods like cross-validation and leave-one-out cross-validation for training the data to find a perfect balance between variance and bias.

References

(N.d.). Retrieved 10 October 2023, from http://zheng-lab.cecs.anu.edu.au/Project/project_prw.html

Pedestrian detection database. (n.d.). Retrieved 10 October 2023, from https://www.cis.upenn.edu/~jshi/ped_html/

Zhang, S., Xie, Y., Wan, J., Xia, H., Li, S. Z., & Guo, G. (2019). WiderPerson: A diverse dataset for dense pedestrian detection in the wild. Retrieved from http://www.cbsr.ia.ac.cn/users/sfzhang/WiderPerson/