CSE343

Intelligent Object Detection and Adaptive Navigation for Autonomous Vehicles using Machine Learning

Shivam Yadav 2020126

Pratham Singhal 2020347

Nitesh Garg 2021075

Divya Raj Singh 2021528

Abstract

The development of self-driving vehicles holds great promise in enhancing transportation safety, efficiency, and convenience. However, it presents substantial challenges in accurately detecting objects and making intelligent decisions within complex and dynamic road environments. This project aims to pioneer a cutting-edge system capable of recognizing various elements on the road, including objects, animals, and pedestrians. It aims to enable informed decision-making for safe and efficient navigation in diverse scenarios, such as encounters with vehicles traveling at specific speeds or unexpected obstacles like a flying ball. Notably, the project is tailored to address the unique challenges posed by Indian traffic and infrastructure, distinguishing it from efforts in other countries. The project strives to achieve superior accuracy, speed, and robustness in pedestrian detection and intent forecasting by leveraging a combination of deep learning models, including YOLO, Faster R-CNN, and HOG + SVM. This report comprises a problem statement, dataset definition, methodologies, results, and project analysis.

1. Introduction

In many economically challenged regions, the intricacies of road conditions present formidable challenges for existing self-driving car models. These models often struggle to adapt to diverse variables such as pose, scale, illumination, and background variations, limiting their applicability to various scenarios. Recognizing the pressing need for improvements, this project seeks to advance current models by incorporating additional variables, enabling them to make more accurate predictions.

The emergence of self-driving vehicles holds immense potential for elevating transportation safety, efficiency, and convenience. However, these vehicles encounter substantial hurdles, particularly in the realms of object detection and intelligent decision-making within intricate and dynamic road environments. This project endeavors to pioneer an innovative system capable of proficiently identifying objects, animals, and pedestrians on the road. Its ultimate goal is to facilitate intelligent decision-making for safe and effective navigation in diverse scenarios, such as encountering vehicles with specific speeds or unforeseen obstacles like a flying ball approaching the vehicle.

Crucially, this project is tailored to the unique challenges posed by *Indian traffic* and infrastructure, which often differ significantly from those in other countries. The focus lies on comprehensive model training to address the complexities of Indian scenarios, where current benchmarks may fall short, leading to errors and misjudgments. The project's primary emphasis is on augmenting holistic object detection and informed decision-making capabilities for self-driving vehicles, recognizing the necessity for a more calibrated approach in the vast and varied landscape of Indian road scenarios.

2. Literature review

2.1. 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.

2.2.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.

2.3.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 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.

3. Dataset

3.1.Data Collection

To expedite data acquisition and significantly enhance data volume. While access to some datasets was restricted due to registration requirements or pending access, or some websites encountered server error issues, we were able to gather data from the following websites for our intended use case:

1. Zheng et al. Dataset: http://zheng-lab.cecs.anu.edu.au/Project/project_prw.html: 11,816 images
2. UPenn Pedestrians: https://www.cis.upenn.edu/~jshi/ped_html/: 170 images with pedestrian size between 180 and 390 pixels
3. WiderPerson: http://www.cbsr.ia.ac.cn/users/sfzhang/WiderPerson/: 13,382 images

Although these datasets were initially designed for pedestrian detection, they can serve our purpose of traffic detection until we gain access to a more suitable dataset or gather our data. This will provide a stepping stone for detecting and developing models before applying the methods to a large dataset. It will also enable us to identify the most appropriate model for our project.

3.2.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 they are running. The model also generates confidence scores for each detection, indicating the likelihood that the enclosed object is indeed traffic. These outputs provide valuable information for downstream tasks like traffic monitoring and analysis.

3.3.Pre-processing Steps

1) Removing Outliers
2) Importing the 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

4. Methodology

4.1.Models Used for Initial Testing

- Histograms of Oriented Gradients
- Support Vector Machines
- Logistic regression
- Gaussian Naive Bayes
- Random Forest

4.2.Model Details

- Logistic Regression:
  - ➔ We used this model to predict whether there is any object in the front of the vehicle.
  - ➔ 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
  - ➔ 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 the 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.

For this project's scope, we decided to move further with the Convolutional Neural Network (CNN) model for various reasons. The shift to CNNs is motivated by their ability to automatically learn hierarchical features, handle spatial complexities, provide scale and pose invariance, facilitate end-to-end learning, adapt to image data, and deliver strong performance on large datasets. These attributes make CNNs a robust choice for addressing the challenges in object detection and decision-making in the dynamic and varied road environments targeted by this project. The same was confirmed by the accuracy and precision of the various models.

5. Results And Analysis

5.1. Findings

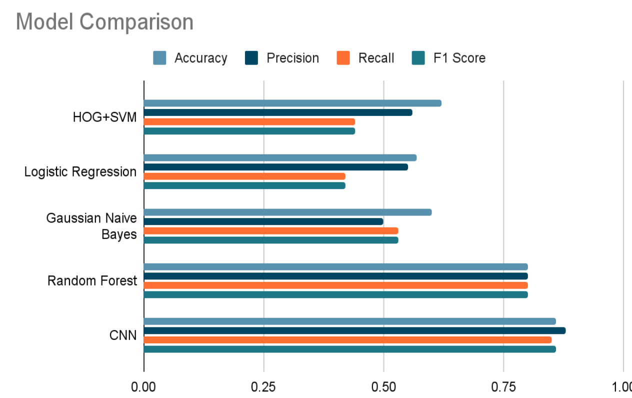
Model	Accuracy	Precision	Recall	F1 Score
HOG+ SVM	62.38%	0.56	0.44	0.44
Logistic Regression	57.61%	0.55	0.42	0.42
Gaussian Naive Bayes	60.02%	0.50	0.53	0.53
Random Forest	83.8%	0.80	0.80	0.80
CNN	86.1%	0.88	0.85	0.86

**This was recorded after tuning the hyper-parameter for some models*

5.2. General Observations

Dataset Characteristics: The decision to transition to CNN is reinforced by the diverse and dynamic nature of Indian traffic scenarios, where the adaptability and learning capabilities of CNNs are expected to outperform traditional methods.

Scalability: CNNs also offer scalability advantages, demonstrating robust performance even as the complexity of the dataset increases.



5.3. 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
4. CNN:
   - Advantages: The decision to transition to Convolutional Neural Networks (CNNs) stems from their innate ability to learn hierarchical features from raw data autonomously. This facilitates handling complex variations in scale, pose, and illumination, making CNNs well-suited for image-related tasks. Additionally, CNNs enable end-to-end learning, eliminating the need for extensive manual feature engineering and providing adaptability to diverse visual scenarios.
   - Performance: In our experiments, the CNN model demonstrated superior performance, achieving higher accuracy, precision, recall, and F1 score than traditional

models. This underscores the effectiveness of CNNs in capturing intricate patterns and representations crucial for object detection and decision-making in dynamic road environments.

6. Conclusion

6.1. 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 to using CNN.
- It helped us to understand the optimisation process of the model to have better results.
- Model can do much better with brake as another possible output for the vehicle as the current model is finding it hard to make sound decisions in case of traffic or overlapping obstructions.
- Model is currently missing features that simplifies the image based on colour, which could save computational time and give faster decisions for maneuvering the car.

6.2. 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.
- Using more outputs like *brake* for the vehicle
- Keeping in mind the speed of the road obstructions could also be kept in mind.
- Object direction could also be explored to make better turning decisions.

In conclusion, the journey so far has equipped us with a deep understanding of machine learning models and their application to autonomous navigation. With a foundation built on diverse models and comprehensive evaluation metrics, we look forward to the future tasks that will propel our project into new frontiers, ensuring its continued relevance and effectiveness in the ever-evolving landscape of autonomous vehicle technology.

References

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

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

3. 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