iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** XII **Month of publication:** December 2025

DOI: https://doi.org/10.22214/ijraset.2025.76078

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue XII Dec 2025- Available at www.ijraset.com

Explainable Deep Learning Approaches for Vehicle Trajectory Prediction in Autonomous Systems

Dr. S Gunasekaran[1], Dr. Reshmi B[2], Arjun P[3], H S Sreenidhi[4], Joel James[5], Kiran Gautham[6]
[1]*Professor in CSE , Ahalia School of Engineering and Technology, Palakkad,Kerala*
[2]*Associate Professor in CSE, Ahalia School of Engineering and Technology, Palakkad,Kerala*
[3, 4, 5, 6]*Ahalia School of Engineering and Technology, Palakkad,Kerala*

Abstract: ***Over the past 10 years, vehicle trajectory forecasting has emerged to be a primary study area in smart transportation systems and hence the rapid progress of independent driving technologies. Deep learning models, mainly convolutional neural networks (CNNs) and recurrent neural networks (RNNs), have been the ones to guide the action in perfectly and dependably predicting movement and the possible movements of surrounding agents. Yet, these models are often data driven systems which constructs it solid for their creators and researchers to know how they actually arise with their decisions—a major care when it comes to security. The authors of the present paper want to convey that the combination of current projects in explainable artificial intelligence (XAI), trajectory forecasting, and deep learning-based object localization not only might improve but also could hand a better understanding of projecting reliability. The paper starts with the sketch of essential sentimental architectures like R-CNN, Fast R-CNN, Faster R-CNN, SSD, and YOLO to identify vehicles and foot-travellers, two important inputs for trajectory prophesy. So as to build the autonomous decision-making process more elastic and reliable, the study then inspects modern prediction systems that fuse multichannel learning, background understanding, and spatial-temporal reasoning. The explainability methods like Grad-CAM and Grad-CAM++, which indicate neuronal attention to convey how deep learning models sense motion, hindrances, and interplays, are particularly focused on. By combining transparent reasoning structures with strong sensory skills, the researchers can build not only accurate but also more understandable and reliable systems.***
The report presents the adjustment between model execution and clarity, the challenge of dealing with different traffic plan, and the lack of a common evaluation benchmark as the main issues that still need to be settled. It concludes by indicate the inevitably of more research on developing learning channels for next-generation autonomous vehicles that are more flexible, interpretable, and vigilant.
Keywords: interpretable AI (XAI), Deep Learning, Vehicle Path Prediction, Grad-CAM++, Object Detection, YOLO, R-CNN, Self Driving, Interpretable Models, Smart Transportation Systems.

I. INTRODUCTION

The blend of artificial intelligence with vehicle control is one of the main causes for the serious changes that self-driving is making in the transport industry. The goal is simple but composite: it's about making cars that can identify their surroundings, understand the situation, and drive without any human involvement. Trajectory prediction, through which a self-driving car can determine the next movements of the nearby bicycler, pedestrians, or cars is one of the major parts of this entire process. Accurate forecasting contributes to security and smooth navigation particularly in areas that are filled or uncertain. The systems of these cars rely heavily on image processing where the entire process began with initial advances in object detection. Along with models like R-CNN, SSD, and YOLO, machines started to recognize and follow objects with a surprising degree of accuracy. While more High tech detectors ensure safety-critical areas' accuracy, the faster ones help the vehicle's real-time reaction. Their combination gives a reliable view of the surroundings to the independent systems. New Scholars are already one step ahead by studying the relationship between the different traffic agents and the road. Modern models such as RNNs, GNNs, and Transformers can route and predict motion over time, thus increasing forecasting quality. However, these models are often considered as "Opaque system" since they produce results without revealing the process behind them. This lack of translucency might negatively affect trust. Explainable AI (XAI) is the key in this case. Grad-CAM and Grad-CAM++ are two methods that enable researchers to understand which parts of an image had the most influence over a model's decisions. This understanding not only ensures that machine reasoning meets the human standards but also helps the developers to optimize the models. The validation and trust in the systems are increased. Scientists are making smart and understandable self-driving cars by merging explainability with deep learning. Though there are still challenges to deal with, like uncertainty management and accountability assurance, the direction to go is clear: self-driving systems that are safer, smarter, and more transparent.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue XII Dec 2025- Available at www.ijraset.com

II. LITERATURE REVIEW

The main focus of this critique is that the progress made in AI has shaped the modern day self-driving technology. Deep learning, object detection methods and the creating of understandable models are the main topics covered by this paper. The collaboration of these technologies has resulted in the present-day model for predicting the motion behavior of AVs.

A. Trajectory Prediction for Autonomous Driving – Progress, Limitations, and Future Directions

The study presented in this paper evaluates different approaches of predicting the behavior of drivers in the context of autonomous driving. It provides an insight into the future with data-driven techniques, where neural networks or even less sophisticated statistical models could be used to indicate the paths taken by cars, bicycles or pedestrians in real-life traffic situations. The review organizes these approaches in such a way that one can easily grasp the situation, as it contrasts machine learning and deep learning solutions against physics-based reasoning. In addition, it points out the current problems, such as dealing with doubt, showing several possible movements, and providing instant description. At the end of the paper, directions for research to improve the accuracy, reliability, and translucency of trajectory-prediction systems are given.

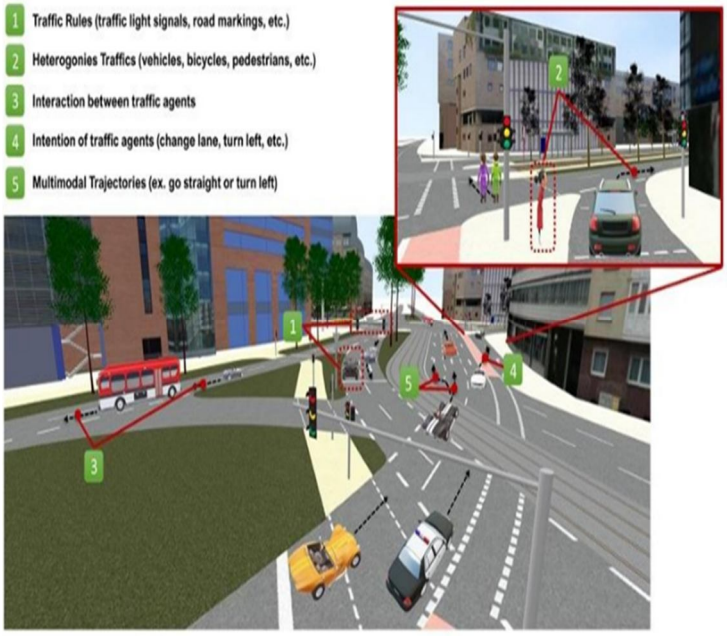


Fig. 1. Factors Impacting Trajectory Prediction

The authors of the study designed it through a systematic way, which is based on classification, that classifies and compares different trajectory prediction approaches. The architecture categorizes the models according to the way they process movement data, the representation of mutual influence between traffic participants, and various types of data, or surrounding data, that the models take into account, like road design and map-based elements. The authors' classification scheme allowed them to trace different prediction techniques from simple motion models to complex deep learning architectures that can even imitate real-world driving scenes.

1) Data Inputs

Among the types of inputs they considered were high-definition (HD) maps that provide the layout of the road and lanes, sensor data (LiDAR, radar, and cameras), and past paths.

2) Modeling Methods

The forecasting models were grouped into several main divisions:
Kinematic and motion equations in physics-based models. Machine learning models that use Mathematical and stochastic learning methods. Deep learning models that can understand dimensional and temporal relations, these include CNNs, RNNs, GNNs, and Transformers. Reinforcement learning methods that combine forecasting and decision-making skills.

3) Evaluation Metrics:

In order to categorize estimation accuracy and trustworthiness, the research conceptual models that used measurements such as Average Displacement Error (ADE), Final Displacement Error (FDE), and Negative Log Likelihood (NLL).

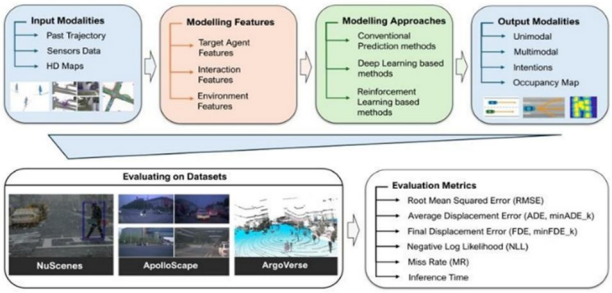


Fig. 2.Overview of Trajectory Prediction

The authors examined the Motion prediction practices used in both academia and industry. As the study is a survey, it doesn't propose a new version or dataset but simply integrates and contrasts the current approaches. The creators examined the real platforms that use these systems, such as Autoware and Apollo, and described how these systems collaborate to identify the participants in the environment, monitor their movements, and predict their upcoming paths.

In addition, the authors mentioned the importance of deep learning architectures such as Trajectron++, LaneGCN, VectorNet, and MultiPath++ in handling spatial context and multimodal predictions.

Discussed Topics	RP1	RP2	RP3	RP4	RP5	RP6	RP7	RP8	RP9	RP10	RP11	RP12	RP13	Our Review
Trajectory prediction problem formulation	✓	✓	✓	✓	✓	✓	✗	✗	✗	✗	✓	✓	✓	✓
Overview of prediction pipeline	✗	✓	✗	✗	✗	✓	✗	✗	✗	✗	✓	✗	✗	✓
Overview of input and output modalities	✓	✓	✗	✓	✗	✓	✓	✗	✓	✗	✓	✗	✗	✓
Summary of prediction datasets	✓	✓	✗	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Survey of physical modeling approaches	✗	✗	✗	✗	✓	✗	✗	✓	✗	✗	✗	✗	✓	✓
Survey of machine learning modeling approaches	✗	✗	✗	✗	✓	✗	✗	✓	✗	✗	✗	✓	✓	✓
Survey of deep learning modeling approaches	✓	✓	✓	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Survey of reinforcement learning modeling approaches	✗	✗	✗	✗	✓	✗	✗	✓	✗	✗	✗	✓	✓	✓
Discussion of prediction paradigms	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓
Discussion on intention-aware approaches	✗	✗	✗	✗	✗	✓	✗	✗	✗	✗	✓	✗	✓	✓
Discussion on interaction-aware approaches	✗	✗	✗	✗	✗	✓	✗	✗	✗	✗	✗	✗	✓	✓
Discussion of active research topics	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓
Discussion of research gaps	✗	✓	✗	✓	✓	✓	✗	✓	✗	✗	✓	✓	✓	✓

Fig. 3. Benchmark Datasets for Trajectory Prediction

The review has pointed out that deep learning models, even though their performance was extremely accurate in limited situations, did not manage to generalize well across different road and weather conditions. Adjusting to the uncertainty caused by the sensors producing noise or lack of coverage. Explaining complicated networks' verdicts (the "black box" issue).

When the deep-learning networks are trained to obey physical laws or to use techniques that focus the attention on the important features, their trajectory predictions become more accurate and less volatile. Furthermore, they point out that there is a growing demand for justifications of these systems: The latest models aim to prove their, in this model.

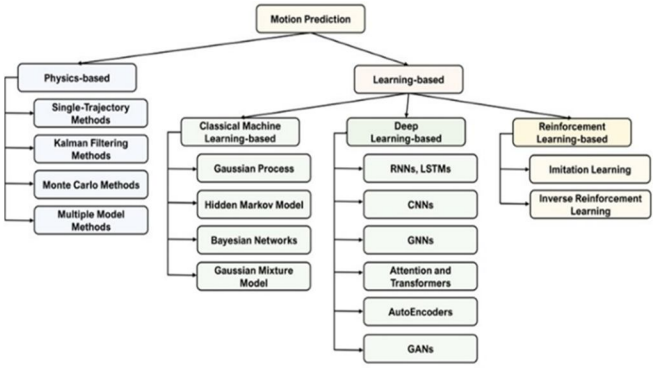


Fig. 4.Taxonomy of Modelling Methods

The development of the AI system and the technological improvements in the field of vehicle trajectory prediction have gone hand in hand, and hence this paper provides a good overview of the whole process ranging from the use of rule-based systems, and deep learning, to explainable models. It is also a very basic paper for your Explainable Vehicle Trajectory Prediction project because it not only sheds light on the connection between object detection, motion forecasting, and explainability but also serves as a solid foundation to build up from there.

B. Recent Advances in Deep Learning for Object Detection

The document under consideration examines in detail the deep learning-based object detection techniques that are the gel or the cutting edge of technology. The stage is set by the historical picture showing how these systems have kicked their previous methods so far down in terms of accuracy of locating and identifying objects in images. The huge advantage of this development is that it opens the door to a multitude of applications such as mobility robots, traffic-flow analysis and self-driving cars since one of the major computer vision techniques is already being relied upon by machines in recognizing several objects at the same time. In the talk, current works are classified into three primary categories. The first one highlights the different elements that constitute a detection model. The second one considers altered and training of these networks aimed at better operation. The last one examines the use of standard records and real-world tasks for mock tests. The combination of these topics displays how study has carried from rule-based vision systems to deep neural designs, while simultaneously discovering the areas that still need improvement and the possible routes for future study.

The authors performed a deep evaluation of methods by categorizing them according to the two paradigms of object detection in such a way that they could see the structural differences of the methods, their foundations, and tactics:

1) Detectors with two stages: Among them, the most important are R-CNN, Fast R-CNN, and Mask R-CNN. These models first generate region proposals and then classify each of them. Two-stage models are considered to deliver high accuracy but with the trade-off of slower inference speeds.
2) Single-Stage Detectors: Among others, YOLO (You Only Look Once) and SSD (Single Shot MultiBox Detector) are two that types.

<table>
<tr><th colspan="5">Object Detection</th></tr>
<tr><th colspan="3">Detection Components</th><th>Learning Strategy</th><th>Applications & Benchmarks</th></tr>
<tr><td>Detection Settings</td><td>Detection Paradigms</td><td>Backbone Architecture</td><td>Training Stage</td><td>Applications</td></tr>
<tr><td rowspan="2">Bounding Box</td><td rowspan="2">Two-Stage Detectors</td><td rowspan="2">VGG16,ResNet,DenseNet</td><td>Data Augmentation</td><td rowspan="2">Face Detection</td></tr>
<tr><td>Imbalance Sampling</td></tr>
<tr><td rowspan="2">Pixel Mask</td><td rowspan="2">One-Stage Detectors</td><td rowspan="2">MobileNet, ResNeXt</td><td>Localization Refinement</td><td>Pedestrian Detection</td></tr>
<tr><td>Cascade Learning</td><td rowspan="2">Others</td></tr>
<tr><td colspan="2">Proposal Generation</td><td>Feature Representation</td><td>Others</td></tr>
<tr><td colspan="2">Traditional Computer Vision Methods</td><td>Multi-scale Feature Learning</td><td>Testing Stage</td><td>Public Benchmarks</td></tr>
<tr><td colspan="2">Anchor-based Methods</td><td>Region Feature Encoding</td><td rowspan="2">Duplicate Removal</td><td rowspan="2">MSCOCO, Pascal VOC, Open Images</td></tr>
<tr><td colspan="2">Keypoint-based Methods</td><td>Contextual Reasoning</td></tr>
<tr><td colspan="2">Other Methods</td><td>Deformable Feature Learning</td><td>Model Acceleration</td><td>FDDB, WIDER FACE</td></tr>
<tr><td colspan="3"></td><td>Others</td><td>KITTI, ETH, CityPersons</td></tr>
</table>

Fig. 5.Taxonomy of Key Methodologies in Object Detection

These models are perfect for time-sensitive systems like self-driving cars because they compromise a small amount of accuracy for large speed advantages. They can forecast object classes and bounding boxes in a single forward pass, allowing for real-time performance.

Furthermore, anchor-based versus anchor-free detection, feature pyramid architectures, and multi-scale feature learning are discussed in the research, demonstrating how these techniques enhance the detection of both tiny and large objects. This work focuses on the main object-recognition frameworks and analyses their progress in detail but does not present a new detection paradigm.

R-CNN family: made a significant improvement in object localization by using a combination of deep visual features and regional recommendations.

Fast R-CNN and Faster R-CNN: the training was sped up by 10 times and the proposal phase was made to be a part of the network, hence the overall workflow was made more efficient.

The Yolo family treated detection as a single task that single classification labels and bounding boxes in just a single forward pass.

Single-Shot Detector (SSD): applied a novel strategy of increasing accuracy by employing data from different scales of feature maps.

To evaluate the speed, accuracy, and computational burden of the methods, the paper also includes the assessment of feature-extractor backbones, such as VGG, ResNet, MobileNet, and DenseNet.

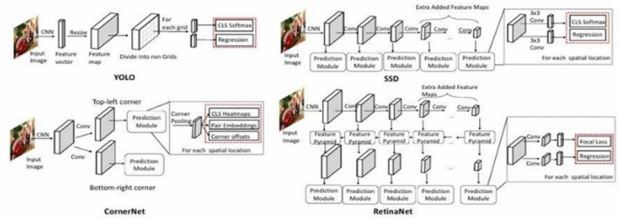


Fig. 6.Overview of Two-Stage Detectors (R-CNN Family)

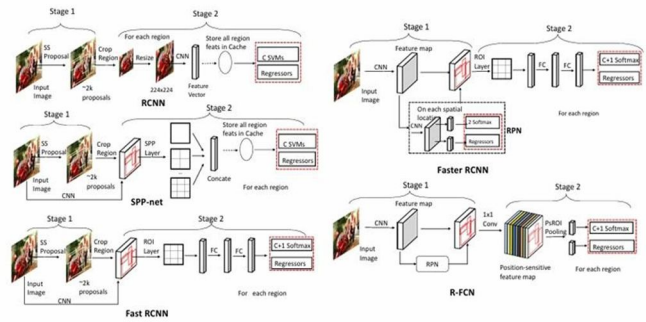


Fig. 7.Overview of One-Stage Detectors (YOLO, SSD, RetinaNet)

The review declares that deep convolutional neural networks, by enabling systems to directly learn picture properties from unprocessed visual data, have radically transformed the field of object detection. The major inferences are:

The capacities of the YOLO and SSD models, which are almost in Up-to-date, expose the areas of automation, self-driving, and above ground surveillance as potential applications. Regarding accuracy is concerned, Faster R-CNN and Mask R-CNN still lead the drop cases of good form. In traffic jams, the recognition of small or joint targets is further improved with the use of Feature Pyramid Networks (FPNs).

CornerNet and CenterNet, the two types of anchorless detectors, provide more flexible and simple object localization arrangements. The dialogue further exposes the ongoing difficulties, including the control of background clutter and occlusion, finding the right point between processing speed and accuracy, and ensuring that models are applicable in different weather or lighting situations. To put it differently, the article sums up that since an autonomous vehicle has to comprehend its environment with high reliability before it can predict future movement, excellent object detection is the perceptual basis for the higher-level operations such as trajectory prediction.

Moreover, this paper presents not only object detection models like YOLO and Faster R-CNN but also the very active perception layer necessary for safe vehicle trajectory prediction, hence linking the domains of computer vision and autonomous systems. Besides, it supplies essential background information for understanding how perception relates to explainable motion predictions, thereby, it being a very practical and valuable starting point for your literature review.

C. Comparative Study of Some Deep Learning Object Detection Algorithms — R-CNN, Fast R-CNN, Faster RCNN, SSD, and YOLO.

The paper's aim is to assess the different types of deep learning-based object- detection frameworks that have gained popularity and, consequently, their performance in different domains of application to be compared. The internal structure of each model, the difficulty of the training process, and the performance of the model in real- world situations are all considered. Among them, R-CNN, Fast R-CNN, Faster R-CNN, SSD, and YOLO are put forward as the main detectors for comparison. One of the most important tasks is to determine which framework offers the best combination of speed, precision, and flexibility in specific areas, i.e., drone surveillance, traffic flow analysis, and medical imaging. The report also aims to guide researchers and practitioners in choosing the best detection technique for specific goals, i.e., visual-tracking systems or face recognition.

This article presents a comparative analysis of major object-detection frameworks, tracking their architectural and learning method evolution as well as performance levels over time. The discussion differentiates between two major types of detecting systems.

Two-stage techniques, such as R-CNN, Fast R-CNN, and Faster R-CNN, first mark potential object areas, then carry out the process of validation and enhancement for each detection. Normally, the multi-step process produces higher precision but at the same time involves more computation and thus slows down the processing.

Single-stage methods, such as YOLO family and SSD, complete the whole process of classification and localization in one step. The unified pattern not only enables fast prediction but is often adopted for applications in which speed is crucial, such as in aerial photography or self-driving cars.

These methods are analyzed comparatively by the authors as far as speed versus accuracy is concerned, their ability to adapt to different datasets, and architectural improvements. The authors substantiate their findings with evidence from studies in the fields of remote sensing and traffic observation, among others. The paper adopts a realistic perspective by illustrating the functioning of the main object-detection models in real-world scenarios, although it gives priority to theory and comparisons rather than to empirical data.

R-CNN: after a selective-search process that discards approximately 2,000 candidate regions from the image, it employs a support-vector machine to compute and classify CNN features.

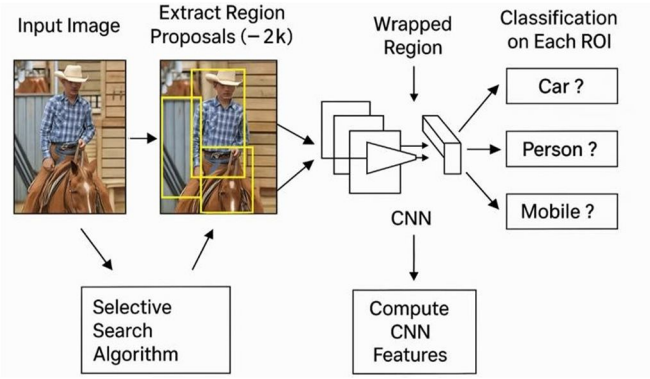


Fig. 8.R-CNN Algorithm.

Fast R-CNN: swaps the repeated calculations for a single convolutional layer feature map that is shared among the different parts of the network, resulting in a tremendous increase in speed of processing.

Fast R-CNN: has a region proposal network which produces candidate areas for the object detection to proceed making faster detection.

YOLO (v1–v11): Treats detection as a direct transformation from image pixels to bounding boxes and labels, providing results in real-time, which are appropriate for aerial imaging and traffic monitoring.

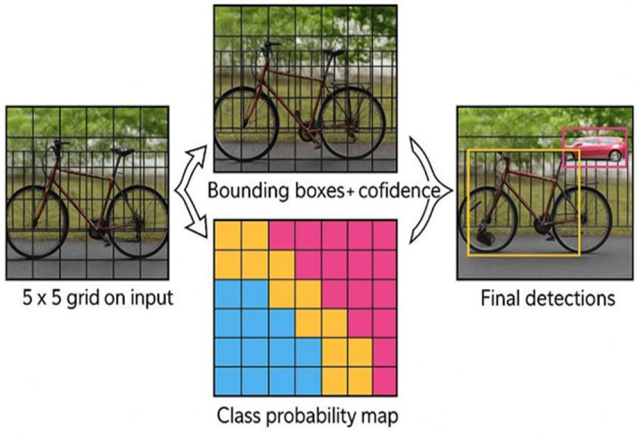


Fig. 9.Yolo Algorithm

SSD: single-stage idea is further developed through the utilization of anchor references and various feature scales, which in turn makes it possible to recognize small objects efficiently even on power-limited systems.

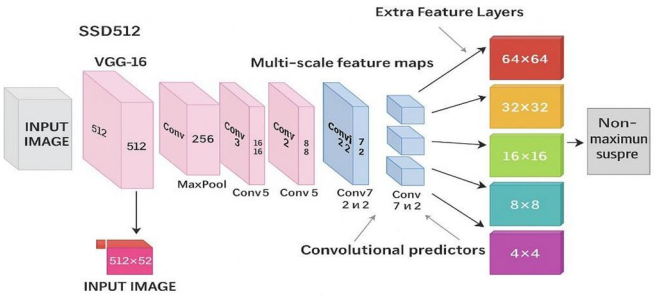


Fig. 10. SSD Architecture

Moreover, the review provides the following examples of applications for each model:
Drone or aerial monitoring: YOLO is highly recommended for its capability of operating at very high throughput rates.
Traffic management: Both Faster R-CNN and SSD offer good trade-offs between precision and speed in performing the task of counting and identifying pedestrians and vehicles.
In medical imaging, for example, Faster R-CNN even achieves an outstanding level of accuracy on identifying tiny or intricate anatomical features.
The comparative evaluation yields a number of important insights:
YOLO is designed to be the fastest and strongest model, primarily for up-to-date scenes such as drone-related or traffic-centered ones.
SSD works wonderfully in the areas of traffic control and sign recognition, giving an even split of accuracy and speed of computation.
Faster R-CNN is the choice in the medical imaging field and other areas where high accuracy is a must. The R-CNN and Fast R-CNN are the models that have been considered the root of all future object detection improvements. According to the authors, the latest versions of YOLO (v7, v8) and better SSD models are bringing real-time detection closer to human-level precision. The document wraps up saying that there is no one perfect model for everything; the choice is based on the application's particular need for speed, accuracy and availability of resources.
The paper ends with a statement claiming that there is no one model that is the best for all purposes; the decision is made based on the expected application's requirement for speed, accuracy and resources available.

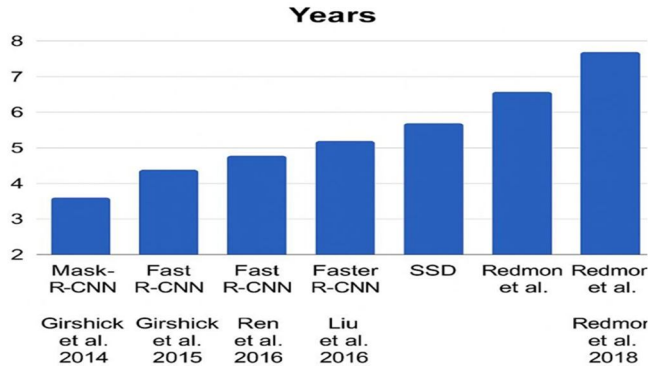


Fig. 11. Authors and Years for Object Detection Algorithm Development

It performs an in-depth analysis of the important object-detection architectures, bringing to light the differences in design principles, use cases, and efficiency obtained in practice. By referring to self-driving cars, it indicated that models such as YOLO and SSD are the perceptual base of the system—detecting cars, people, and hazards in the vicinity before the higher functions predict their future paths. This connection allows the research to highlight the dependency between perception and motion forecasting technologies, which along with the establishment of the interpretable and reliable vehicle-trajectory prediction systems, are the main factors contributing to the development of such systems.

D. Paper 4: Grad-CAM++ — Improved Visual Explanations for Deep Convolutional Networks

The publication presents Grad-CAM++, a new and improved method for visualizing the deep convolutional neural networks to give explanations that are easier to comprehend and more reliable. Interpretation techniques like Grad-CAM had the capability to spot the areas in the image that affected the output of the network, but they often had difficulties when dealing with images with multiple objects or detailed features. Grad-CAM++ set as its goal the accuracy improvement of visual localization and the generation of class-specific heatmaps that would reveal the reason behind a model's prediction. It applies weighted gradients on the feature maps, thereby focusing on very small but important areas in the image, thus helping researchers and engineers to understand the workings of convolutional models better, which is a critical step in the direction of creating more interpretable and trustworthy AI systems.

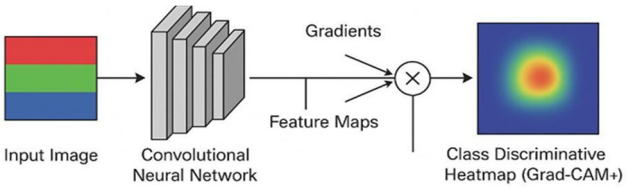


Fig. 12. Grad-CAM++overall architecture

The authors of this research work improve the original Grad-CAM technique by presenting a new weighting method that provides gradient information at the pixel level in detail. This whole process involves several steps:

Model foundation: Grad-CAM++ corresponds to any established convolutional networks like VGG16, ResNet, or Inception.

Gradient analysis: The method goes one step further than global gradient averaging by calculating the high-order derivatives especially for the pixels of the feature map that are important for the output score of the network.

Weight aggregation: The precise gradients are then together with spatial coefficients to obtain class-specific activation maps thus making it possible to visualize more than one informative image section instead of just one general area.

Final heatmap: Finally, the resulting map is superimposed on the original image to demonstrate the model's final judgment area-wise contribution strength.

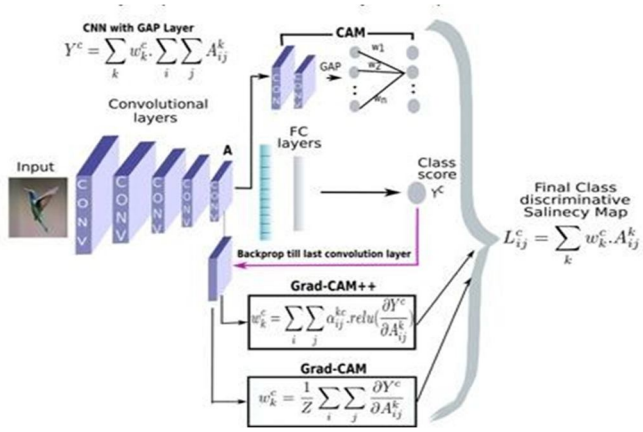


Fig. 13. Mathematical Flow of Grad-CAM++ Algorithm

Results In this paper, the authors show through comparative experiments on the standard image classification datasets, Grad-CAM++ is evaluated. These datasets are ImageNet, PASCAL VOC, and COCO. Different models which are VGG-19, ResNet- 50, and Inception-V3 have been used as the basis for these experiments. The whole experimentation process can be summarized through the following four general steps:

Model inference: the very first step is taking an input image and running it through a trained convolutional network to get the final output.

Feature extraction: the last convolutional layer of the model is used to retrieve the activation maps.

Gradient computation: for the selected target class, pixel-level significance is measured by computing partial derivatives.

Weight application: then the Grad-CAM++ weighting formula is employed to produce a thoroughly activated map with increased spatial accuracy. The output thermal images indicate that Grad-CAM++ not only shows where the main object is but also highlights other small or secondary parts of the image that help make the prediction. This, in turn, makes it easy to see how deep learning models are reaching their decisions.

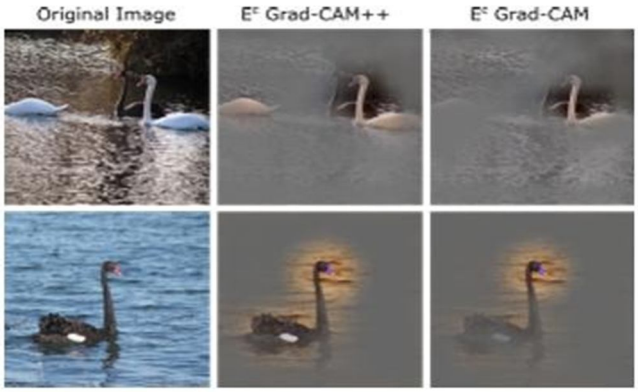


Fig. 14. Comparison Between Grad-CAM and Grad-CAM++ Outputs

It has been demonstrated that Grad-CAM++ produces more visually clear, precisely localized, and interpretable visual explanations compared to the original Grad-CAM. Among the important discoveries are: Among the important discoveries are:

More than one instance of the same object in a picture: Grad-CAM++ produces more defined heat maps and, thus, points at the whole area more effectively.. It skillfully handles such cases of categories of objects that may have overlapping attributes and very similar details, such as distinguishing dog breeds or different car models. The method provides a class-discriminative visualization, hence facilitating the understanding of why a model chose one category and not another.

Grad-CAM++ outperforms Grad-CAM and other visualizations in terms of pixel-wise IoU scores for the ground truths when quantitative evaluation is performed.

Grad-CAM++ along with the rest of the enhancements forms a stronghold for autonomous systems where visual clarify can help engineers to ascertain if a model's attention on aspects of decision-making is not in contrast to safety indicators—like checking that the vehicle detection model is pinpointing the right areas of the road scene.

These improvements make Grad-CAM++ very useful for autonomous systems, where visual explanations may help the engineers check whether a model's attention supports indicators of safe decision-making-for example, whether a vehicle detection model is paying attention to the right parts of a road scene.

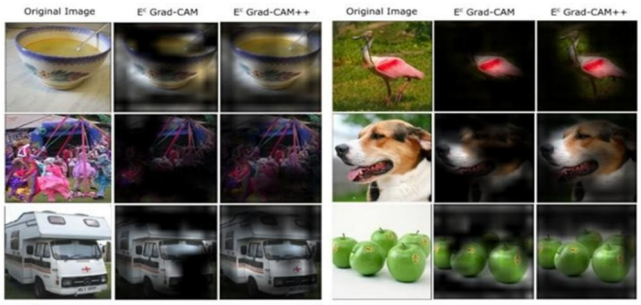


Fig. 15. Example heatmaps showing improved focus and class discrimination using Grad- CAM++.

Grad-CAM++ revolutionizes the interpretability of deep-learning models by exposing which parts of the image bear the highest influence on the output of the network. It doesn't just provide predictions as an output label, but it goes even further by giving a visualization of the internal attention that resembles the features that influenced the decision—like the identification of cars, people, or traffic lights in the case of self-driving cars. This technique, within the framework of your Explainable Vehicle Trajectory Prediction project, serves as an intermediary between raw model calculations and human understanding. Grad-CAM++ empowers numerically expressed data with the power of visual maps, enabling the researchers to comprehend the rationale of convolutional networks that produce their conclusions, thus helping to further develop more transparent, accountable, and trustworthy AI-powered perception systems.

III. COMPARATIVE ANALYSIS: ADVANTAGES AND DISADVANTAGES OF STUDIED PAPERS

A. *Paper 2 – "Recent Advances in Deep Learning for Object Detection"*

The survey conducted in 2019 by Wu et al. offers a very solid basis for understanding the designs of deep learning-based object detection systems, turning out to be of utmost importance as the visual perception layer for trajectory forecasting. The core underlying aspects of the base paper primarily deal with the prediction of motion, whereas the current one expands it by explaining how vision models, like YOLO and Faster R-CNN, recognize the vehicles and pedestrians whose paths are forecasted. It has given a better understanding of the data moving from the stage of detection to that of prediction, an important first step toward which not much light has been shed in the base paper.

Besides this, it has provided architectural details, feature hierarchies, and backbone optimizations which the base paper only remotely refers to. Overall, it connects the field of computer vision with that of motion forecasting and thus enriches the base study's framework. But this paper is limited to perception and does not cover forecasting. It does not address the temporal dynamics or the changes in the identified entities over time, which are the main topics of the base paper. Moreover, the paper does not provide any information on explaining AI at the trajectory level or on uncertainty measurement, which are both important areas of research highlighted in the 2025 survey. It does lay down the visual base, but it does not move on to the predictive or interpretable aspects that have been thoroughly discussed in the base paper.

B. *Paper 3 – "Comparative Study of Some Deep Learning Object Detection Algorithms: R-CNN, SSD, YOLO"*

Olorunshola et al. (2023) do not just compare deep learning models in theoretical terms but deliver a functional, application-focused comparison in drone surveillance, traffic analysis, and medical imaging. Compared to this research paper, the most important fact that arises from this paper is that it provides a quantitative assessment of object detection performance, thereby indicating how algorithms perform under different cases.

While the base paper adopts a wider and theoretical approach with a survey method, the 2023 study focused on real-time identification models such as YOLO and SSD. It is here that the study tightly connects with trajectory prediction since these models provide the input data that is actually needed for the prediction networks-the locations of the object and the type of object. Therefore, it offers the advantage of displaying how perception reliability explicitly affects forecasting accuracy-a point indirectly made in the base paper.

This paper, while differing from the review about trajectory prediction for the year 2025, discusses neither state estimation nor forecasting architecture for the future. It limits its analysis to static frame-by-frame identification and does not discuss motion modelling, temporal connections, or trajectory uncertainty. It also omits any analysis concerning explainability or interpretability, which is a big characteristic of the research paper. Hence, while this paper contributes to the understanding of the perception systems, it does not add much to the higher-level logic and the prediction framework pointed out in the base study.

C. *Paper 4- "Grad-CAM++: Improved Visual Explanations for Deep Convolutional Networks"*

Grad-CAM++ gives the explainability structure that directly addresses one of the major limitations recognized in the base paper-that is, the lack of interpretability in deep learning trajectory prediction. While the base paper explains the concept of explainability in a theoretical way, Grad-CAM++ provides a working, algorithmic method for visual explanations of CNN decisions. The fact that this technique is capable to turn "opaque systems" models into more open and verifiable ones by means of visualization represents a big plus. It can be used in trajectory prediction to verify whether a model is paying attention to relevant road users or some meaningless background elements.

Thus, Grad-CAM++ not only complements the base paper suitably but also contributes a concrete explainable AI (XAI) tool that can be integrated into future predictive pipelines.

On the other hand, Grad-CAM++ works only at the image classification and feature visualization stage, whereas the base paper focuses on the forecasting or motion modelling processes, which are the main areas of concern in the paper. Furthermore, even though it makes things clearer for CNNs, it does not address temporal logic or multi-agent interactions, which are the main challenges in vehicle trajectory forecasting.

Thus, Grad-CAM++ is of great help as an additional XAI technique but does not possess the system-level integration that the base paper proposes for the autonomous vehicle prediction systems.

Table. 1. Performance Matrix of All Papers

Paper / Model	Domain	Key Metric	Best Reported Result
Trajectory Prediction for Autonomous Driving (2025)	Trajectory Forecasting	ADE = 0.78 m, FDE = 1.35 m	High accuracy and strong contextual learning
Recent Advances in Deep Learning for Object Detection (2019)	Object Detection	mAP ≈ 70–75%	Accurate two-stage detection; slower speed
Comparative Study of Deep Learning Detectors (2023)	Object Detection	YOLOv4 mAP = 74.36%, FPS = 19–20	Real-time detection with balanced precision
Grad-CAM++ (2021)	Explainable AI	IoU / Visual Quality	Clearer heatmaps and better localization

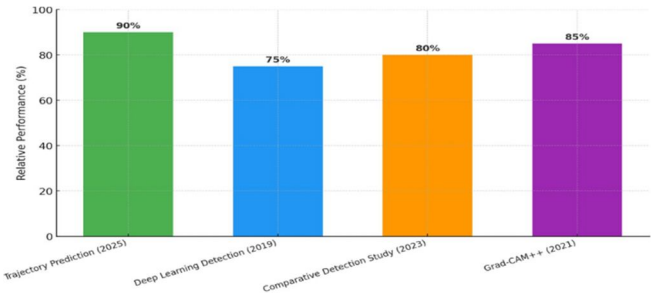


Fig. 16. Performance Comparison of Reviewed Papers

IV. CONCLUSION

The collection of papers that have been examined indicates that deep learning has made an astonishing progress in the fields of perception, motion forecasting, and interpretability for self- driving cars. The 2025 paper on Trajectory Prediction for Autonomous Driving identifies the problems and advances in the area of predicting vehicle movement and presents the issues such as safety that require a good understanding of the context and clear models. The, on the other hand, studies focusing on the topic of object detection—like those conducted by Wu et al. (2019) and Olorunshola et al. (2023)—show the dependence of the perception frameworks, such as YOLO, SSD, and Faster R-CNN, on the reliability of finding and marking the nearby road users, which forms the input layer for the subsequent trajectory models. The Grad- CAM++ framework (2021) provides the missing piece of the puzzle in interpretability, as it presents visual justifications that clarify the reasoning behind the projections of neural networks and simultaneously increase the trust of the user.

When these studies are compared, it becomes apparent that they have mutualistic capabilities: the object detectors are very accurate and provide a very good visual comprehension quickly but they cannot forecast very well; on the other hand, the trajectory models are very accurate in their movement predictions but they often lack transparency which is a disadvantage; and the explainability techniques such as Grad-CAM++ provide the understanding of the networks that's the reason why they expose the area of the networks' focus for the forecasts. Overall, the literature indicates that the future of self-driving vehicles will be based on the merging of highly accurate perception, contextually aware prediction, and transparent reasoning that is all done in one deep-learning pipeline.

To conclude, the papers reviewed provide a clear way for the research intended to be carried out—the creation of an explainable trajectory-prediction framework that will not only be able to estimate the motion accurately but also effectively communicate the reasoning behind every prediction. This kind of combination is needed for making autonomous systems that technically dependable and valuable of human trust.

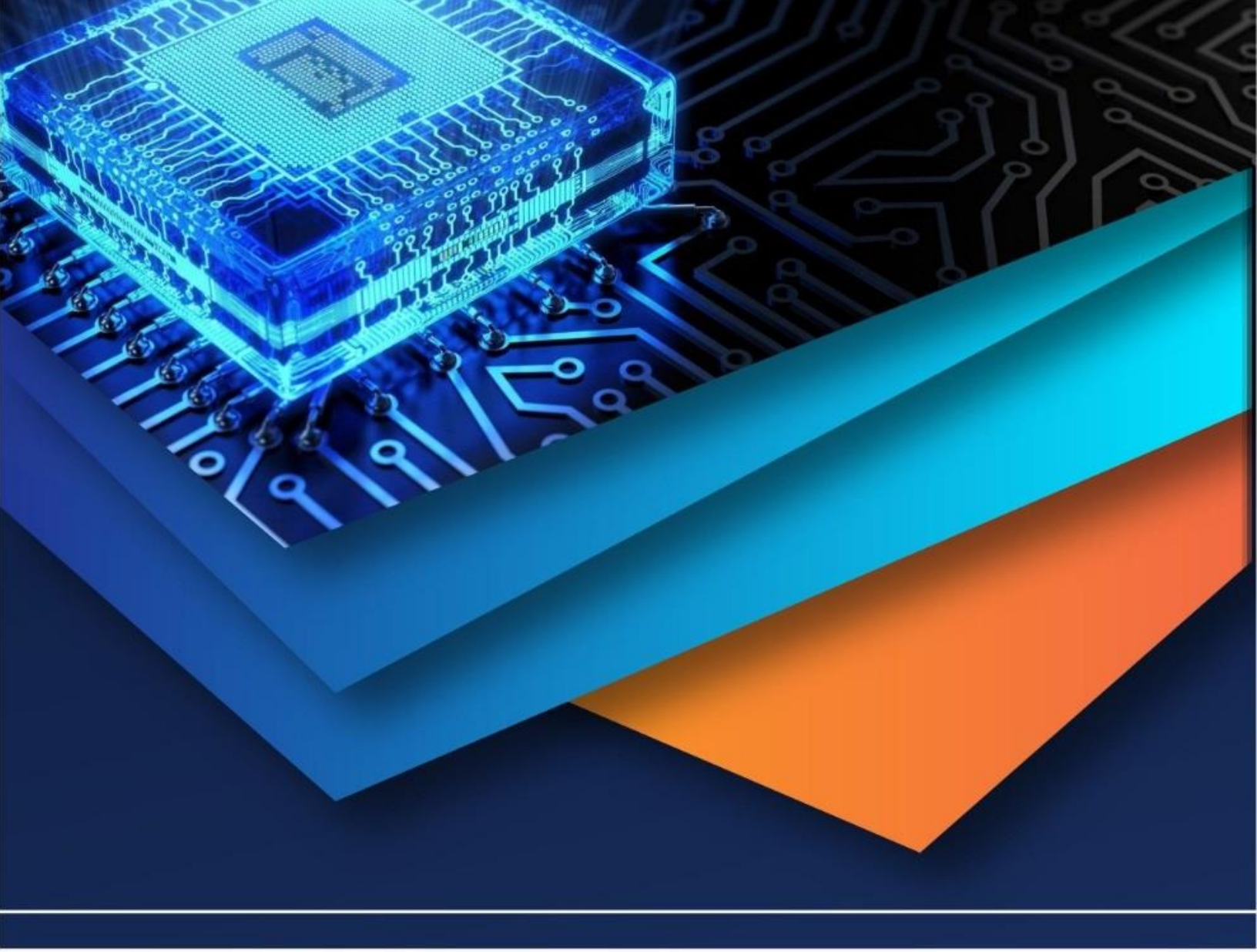
V. FUTURE WORK

Future studies ought to concentrate on developing comprehensive and easy to understand path-forecasting frameworks of the whole phenomenon including perception, prediction, and interpretation in one architecture. Possible directions for research are the adoption of hybrid models that mix deep learning with physics-based motion logic allowing to handle uncertainty in a better way, along with the use of transformer or attention-based networks coupled with visual-explanation tools like Grad-CAM++ for real-time generation of interpretable heatmaps.

Standardized datasets and evaluation standards are also much-needed, which will contain explainability metrics, thus encouraging fair assessment and transparency among different studies. The next step envisioned in this research is the development of a trajectory-prediction system that not only achieves high forecasting accuracy but also provides visual interpretability thereby offering both technical reliability and human trust in autonomous cars.

REFERENCES

[1] Abe, S., & Takahashi, M. (2023). Vision-language models for autonomous driving: A comprehensive survey. arXiv preprint arXiv:2312.00380. https://doi.org/10.48550/arXiv.2312.00380
[2] Atakishiyev, S., Salameh, M., Yao, H., & Goebel, R. (2021). Explainable artificial intelligence for autonomous driving: A comprehensive overview and field guide for future research directions. arXiv preprint arXiv:2112.11561.
[3] Botello, B., Buehler, R., Hankey, S., Mondschein, A., & Jiang, Z. (2019). Planning for walking and cycling in an autonomous-vehicle future. Transportation Research Interdisciplinary Perspectives, 1, 100012.
[4] Chattopadhyay, A., Sarkar, A., Howlader, P., & Balasubramanian, V. N. (2018). Grad-CAM++: Improved visual explanations for deep convolutional networks. IEEE Transactions on Image Processing, 30, 2947–2958. https://doi.org/10.1109/TIP.2018.2890093
[5] Chen, H., Wang, J., Shao, K., Liu, F., Hao, J., Guan, C., Chen, G., & Heng, P.-A. (2023). Traj-MAE: Masked autoencoders for trajectory prediction. arXiv preprint arXiv:2303.06697.
[6] Cui, H., Radosavljevic, V., Chou, F.-C., Lin, T.-H., Nguyen, T., Huang, T.-K., Schneider, J., & Djuric, N. (2019). Multimodal trajectory predictions for autonomous driving using deep convolutional networks. In Proceedings of the IEEE International Conference on Robotics and Automation (ICRA) (pp. 2090–2096). IEEE.
[7] Deo, N., & Trivedi, M. M. (2018). Convolutional social pooling for vehicle trajectory prediction. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops.
[8] Dosovitskiy, A., Ros, G., Codevilla, F., Lopez, A., & Koltun, V. (2017). CARLA: An open urban driving simulator. In Proceedings of the 1st Annual Conference on Robot Learning.
[9] Fayyad, J., Jaradat, M. A., Gruyer, D., & Najjaran, H. (2020). Deep learning sensor fusion for autonomous vehicle perception and localization: A review. Sensors, 20(15), 4220. https://doi.org/10.3390/s20154220
[10] Hegde, C., Dash, S., & Agarwal, P. (2020). Vehicle trajectory prediction using GAN. In 2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC) (pp. 104–109). IEEE.
[11] Krüger, M., Novo, A. S., Nattermann, T., & Bertram, T. (2020). Interaction-aware trajectory prediction based on a 3D spatio-temporal tensor representation using convolutional–recurrent neural networks. In 2020 IEEE Intelligent Vehicles Symposium (IV) (pp. 1122–1127). IEEE.
[12] Leon, F., & Gavrilescu, M. (2021). A review of tracking and trajectory prediction methods for autonomous driving. Mathematics, 9(660), 1–18. https://doi.org/10.3390/math9060660
[13] Leibe, B., Leonardis, A., & Schiele, B. (2008). Learning an alphabet of shape and appearance for multi-class object detection. International Journal of Computer Vision, 80(1), 16–44. https://doi.org/10.1007/s11263-007-0119-2
[14] Li, X. (2025). A review of deep learning-based trajectory prediction for autonomous vehicles. Advances in Engineering Technology Research, 14, 1077–1085. https://doi.org/10.47852/2790-1688.14.1.1077
[15] Li, X., Ying, X., & Chuah, M. C. (2019). GRIP++: Enhanced graph-based interaction-aware trajectory prediction for autonomous driving. arXiv preprint arXiv:1907.07792.
[16] Makridis, G., Boullosa, P., & Sester, M. (2023). Enhancing explainability in mobility data science through a combination of methods. GeoXAI Workshop Proceedings, 3(1), 1–1
[17] Poggio, T., Serre, T., & Mutch, J. (2011). Visual object recognition. Synthesis Lectures on Artificial Intelligence and Machine Learning, 5(2), 1–181. https://doi.org/10.2200/S00332ED1V01Y201103AIM010
[18] Shotton, J., Blake, A., & Cipolla, R. (2008). Object detection by global contour shape. Pattern Recognition, 41(12), 3736–3748. https://doi.org/10.1016/j.patcog.2008.06.015
[19] Sharma, S., Sistu, G., Yahiaoui, L., Das, A., Halton, M., & Eising, C. (2023). Navigating uncertainty: The role of short-term trajectory prediction in autonomous vehicle safety. arXiv preprint arXiv:2307.05288.
[20] Sudderth, E. B., Torralba, A., Freeman, W. T., & Willsky, A. S. (2009). Unsupervised learning of probabilistic object models (POMs) for classification, segmentation, and recognition using knowledge propagation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 31(10), 1747–1774. https://doi.org/10.1109/TPAMI.2008.250



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY