

Final Project Research Paper

Deep Learning for Self-Driving Cars (Autonomous Driving)

1. My Deep Learning Application :

There are many possible deep learning applications, such as medical image analysis, language translation, chatbots, fraud detection, and recommendation systems. But I chose self-driving cars (autonomous driving) because:

1. It is very real and current. Companies like Waymo, Tesla, Uber, Mercedes-Benz and others are actively testing or deploying self-driving systems on public roads.
2. Almost everyone has heard about self-driving cars, so it is easy to relate to in daily life.
3. It is a complex application that uses many deep learning techniques together :
 - Computer vision (seeing the road, objects, lanes)
 - 3D perception (using LiDAR and radar)
 - Time-series prediction (predicting how other cars and people move)
 - Planning and control (deciding steering, acceleration, and braking)
 - Sensor fusion (combining camera, LiDAR, radar, GPS, maps)

There are many recent research surveys and papers that I can use to understand the state of the art and cite in this report.

So this application allows me to clearly show how deep learning is used in real-world systems and what the open research problems are.

2. Background: How Self-Driving Cars Work at a High Level :

Self-driving cars are also called **autonomous vehicles (AVs)**. The standard way to describe their software is to split it into several big blocks or “modules”:

1. **Perception** – “What is around me?”
2. **Localization and mapping** – “Where am I on the map?”
3. **Prediction** – “What will other road users do next?”
4. **Planning** – “What should my car do next?”
5. **Control** – “How do I physically turn the steering wheel, brake, or accelerate?”

Deep learning is heavily used in **perception**, **prediction**, and more recently also in **planning and control** through end-to-end models.

2.1 Sensors Used by Self-Driving Cars

A self-driving car reads the world using several sensors at the same time:

- **Cameras** – give 2D color images; similar to human eyes.
- **LiDAR** – gives 3D point clouds (distance measurements around the car).
- **Radar** – measures distance and speed of objects, especially useful in bad weather.
- **Ultrasonic sensors** – short-range sensors for parking and very close obstacles.
- **GPS & IMU** – global position and motion (acceleration, rotation).

Deep learning models often take **camera images**, **LiDAR point clouds**, or a combination (sensor fusion) as input.

3. Literature Review - Deep Learning Techniques in Self-Driving Cars :

This section is a comprehensive survey of the selected deep learning methods used in autonomous vehicles. The goal of this section is to provide a comprehensive overview of existing deep learning methods, their purpose and usefulness, and a high-level overview of how those existing methods work in everyday terms.

The literature review has been organised based on the these major sub-tasks:

1. Object Detection and Classification
2. Lane Detection and Road Segmentation
3. 3D Object Detection with LiDAR

4. Multi-speaker Fusion
5. Trajectory Prediction
6. Planning and Control via Deep Learning
7. End-to-End Autonomous Driving

3.1 Object Detection and Classification :

Issue: A car must continuously answer the question "What can I see around me?" when answering the question of what it sees. Those items include:

- a. Car
- b. Truck/Buses
- c. Motorcycle
- d. Bicycle
- e. Pedestrians
- f. Traffic Lights
- g. Traffic Signs
- h. Cones/ Barriers/ Animal etc.

To assist with this, deep learning is being used to accomplish it via object detection networks and image classification networks.

3.1.1. Convolutional Neural Networks (CNN)

Most object detection in self-driving vehicles uses CNNs. A CNN takes an image and "learns" filters (or "small windows") that will slide across an image to "find" various "patterns" in that image, e.g. edges, corners, shapes, and increasingly complicated items like "wheel" and "car."

There have been several early versions of CNNs used within many of the same types of applications as well as later on in object detection structures such as:

- Faster R-CNN
- SSD(Single Shot Detector)
- YOLO(You Only Look Once)

The output of those models are typically bounding boxes and labels (for example, "This is the car," "This is a person") with a confidence score assigned.

The reason CNNs are used for this task,

- CNNs make image processing very efficient,
- CNNs can learn from data without needing manual rule creation.
- They can reach high accuracy on standard driving datasets (e.g., KITTI, COCO, Waymo Open Dataset).

How YOLO-style models work :

- The image is divided into a grid (for example 13×13 cells).
- For each cell, the network predicts a few possible bounding boxes and class probabilities.
- The network is trained end-to-end so that these predictions match the ground-truth boxes.
- At inference time, the network gives very fast predictions, which is important for real-time driving.

Recent object detection surveys show deep learning approaches are now the standard way to detect vehicles and pedestrians in autonomous driving.

3.1.2 Transformer-Based Detection

Recently, **Transformers** have been introduced for object detection, for example **DETR (DEtection TRansformer)**. A survey on vision transformers in autonomous driving explains that transformers can capture global context in the image better than CNNs, because each “token” can attend to every other token.

In simple words:

- A transformer looks at the whole image at once, not just local windows.
- It can better understand relationships like “this small object in front of the truck might be a child crossing the road”.

As a result, transformer-based detectors can improve detection in complex scenes, especially in crowded or cluttered environments.

3.2 Lane Detection and Road Segmentation

The existence of several methods in which the car can determine the status of the road, including the state of the lane, where, when, and to what degree can it safely be driven.

In general, segmentation tasks are completed using a Family of Models known as Semantic Segmentation Models. Semantic Segmentation defines, for each pixel in an image, the Network will classify what that pixel represents (Road, Lane, Vehicle, Sidewalk, Grass...).

3.2.1 UNet & Encoder / Decoder Architecture.

UNet is the representative model when thinking of a Segmentation model. The UNet Model contains:

- a. Encoder (compressing the complete image function down to Coarse Feature maps)
- b. Decoder (which expands those feature maps back out to the Original image size producing a pixelwise labeled map).
- c. There are also Skip Connections between the Encoder and Decoder networks, preserving the finer details of the Image.

While these characteristics of the UNet model ultimately benefit Lane Detection models because:

- a. Lane Markings are Long & Thin
- b. Because the model requires Local Detail (Sharp Edges) and Global Structure (Lane Curvature)

Thus, the Encoder / Decoder CNN architecture with Skip Connections captures both.

Other Lane Detection Models include SCNN, SegNet, LaneNet & overall as shown in recent surveys performed evaluating Autonomous Driving Perception, the accuracy of Segmenting the Lane or Road is typically 95% or greater in good weather, with significant reductions in Performance during inclement weather, poor lighting conditions, and faded markings or lane lines.

3.2.2 - Segmentation - A Key Tool for Developing a Driving Plan

The planner can now do the following with the knowledge of the "drivable area," via pixel information:

- Avoid driving onto sidewalks and grass
- Stay correctly within the lane
- Define lane merging/splitting areas

Therefore, segmentation is more than just a visual aid; it also provides valuable data that will assist in creating driving plans, as noted above.

3.3 3D Object Detection with LiDAR

Cameras give beautiful images but do not directly measure depth. LiDAR (Light Detection and Ranging) sends laser pulses and measures their return times, creating a **3D point cloud** around the car.

Problem: The car needs to understand this 3D cloud and detect objects in it.

Recent surveys review deep learning methods for 3D object detection in autonomous driving, which include

- **PointNet / PointNet++** – operate directly on point clouds.
- **VoxelNet** – convert points into small 3D boxes (voxels) and apply 3D CNNs.
- **PointPillars** – compress 3D into vertical “pillars” and then apply 2D CNNs in bird’s eye view.
- **Transformer-based 3D detectors** – e.g., methods reviewed in transformer surveys.

Why 3D detection is needed:

- It gives accurate distance and relative speed of objects.
- It is more robust in low-light conditions than cameras.
- It allows safe path planning around other cars, pedestrians, and obstacles.

However, LiDAR sensors are expensive, produce large data volumes, and their point clouds are sparse. Deep learning models must be efficient enough to process these in real time.

3.4 Sensor Fusion (Camera + LiDAR + Radar)

One sensor cannot perform perfectly because:

- Camera -- color is good, resolution is good; darkness and fog, performance is bad
- LiDAR -- distance measurement is accurate, but cost is high, some areas lack density
- Radar -- can pierce through weather, but resolution is bad

Therefore, self-driving vehicles use sensor fusion. A recent review of deep learning applied to autonomous vehicle and also another survey regarding fusion methodologies includes an explanation of fusing data at various levels:

- **Early Fusion** -- Raw data (all sensor data) are fused together prior to feature extraction.

- Mid-Level Fusion -- Each sensor's data is processed by its own network to generate its own feature map; the two networks' feature maps are then fused (merged) together.
- Late Fusion -- After making predictions using different networks, predictions from each sensor are combined to form the final prediction.

Recent research suggests performing deep fusion at the feature level since:

- It enables the model to determine "where to trust which sensor."
- An example would be trusting a camera for color traffic signs; LiDAR for depth; and radar for low visibility.

A modern idea that has emerged is called BEV (Bird's Eye View), which is employed by methods such as BEVFormer. This technique utilizes transformer architecture to create a single representation of multiple views (from multiple cameras or in some cases, from LiDAR), and then detect and segment objects within that view.

This representation is easier for planning because it looks like a map, not a perspective image.

3.5 Trajectory Prediction

In anticipating the future actions of other roadway users, cars create predictions of the likely actions of other roadway users, for example:

- Will I see a pedestrian cross the street?
- Will the car next to me on the street merge into my lane?
- Will I see a cyclist turn to their left at an intersection?

Trajectory prediction is the anticipated location of agents over a short period of time (approximately 3 to 5 seconds in the near future). Research indicates that trajectory prediction is integral to ensure that autonomous cars can operate safely. The ability to predict a future trajectory of agents also may reduce the likelihood an accident resulting from a poor prediction.

3.5.1 Methods of Predicting Future Trajectory's from Deep Learning

There are various types of neural networks used to predict future trajectory:

- **RNNs/LSTMs/GRUs** - model sequences of historical positions and predict future positions.
- **Graph neural networks** - modelling of one agent as a node (the position of the vehicle) and their interactions with other agents at a given time point as an edge (e.g., other vehicles located in the same intersection).

- **Transformers** - self-attention to learn complex interactions among agents located within an operational area.

These types of networks typically leverage:

- History, including the location of agents over a period of time.
- High-definition (HD) Maps.
- Perception outputs from surrounding environment.
- Context surrounding a scene (crosswalks, traffic lights, etc.).

Research shows that deep learning models provide a clear and significant state-of-the-art improvement when compared to traditional trajectory prediction models (for example, Kalman filter) on the multi-agent trajectory prediction tasks. However, traditional trajectory prediction methods continue to experience challenges dealing with highly interactive and/or rare scenarios (such as unusual pedestrian behaviour or aggressive lane-change behaviour).

3.6 Planning and Control with Deep Learning

The vehicle must now formulate a plan regarding which path it should take, how quickly it should travel, as well as whether it should decelerate or steer, etc., after receiving information both through sensors and predictions about its future position.

Historically, two techniques were utilized to facilitate planning and control in vehicles:

- Rule-Based Systems
- Optimization-Based Planners e.g., Model Predictive Control

Now researchers have begun using Reinforcement Learning (RL) and other types of model-based learning for planning and control in vehicles.

3.6.1 Reinforcement Learning (RL) Applied to Vehicle Driving

In RL systems:

- The agent is the vehicle;
- The vehicle's surroundings are the road network and other vehicles;
- State is defined by the present sensor readings;
- Actions the vehicle can perform include steering, opening the throttle, using the brake system etc.;
- The reward function may include negative values for collisions or for poor comfort, while providing positive feedback for smooth and safe vehicle operation.

When using Deep RL techniques such as DQN, PPO, and SAC within a simulation environment, the vehicle learns how to follow certain techniques when performing driving maneuvers. These deep learning algorithms learn the following driving techniques:

- Maintaining Lane Position
- Overtaking Traffic
- Following Speed Limits
- Executing Unprotected Left Turn Maneuvers

However, using RL for driving in real-world conditions poses several challenges for researchers:

- Learning via trial-and-error processes in the real world is inherently unsafe for vehicles.
- The simulation must accurately reflect the characteristics of the real world in order for the learned action policy to be transferable to real-world conditions.

3.7 End-to-End Autonomous Driving

An alternative to the modular approach is the end-to-end (E2E) approach for Autonomous Driving. With end-to-end autonomous driving, all of the functionality (perception, prediction, planning, control) are implemented in a single unified network instead of being implemented as separate blocks. The end-to-end approach takes in all the raw data from cameras, LiDAR, etc. and directly outputs the commands/trajectories the vehicle should follow.

The end-to-end approach is defined in a recent survey as a fully differentiable program that maps sensor data directly to a driving plan.

The reasons why researchers prefer to use end-to-end models are:

- Simpler architecture and fewer hand-coded rules.
- All parts of the model can be trained together, which allows the model to optimize the entire driving task.
- When sufficient data is available, end-to-end models can learn to perform very complex behaviours.

The reasons why E2E models are controversial are:

- It is more difficult to get an interpretation of what the car "saw", or "how it came to a decision", since it is not easy to separate those two domains.
- End-to-end models are typically much more difficult to debug and have less capability to provide guarantees for safety.
- The internal logic of E2E models is opaque; therefore, regulators may have less confidence in E2E models compared to modelling approaches with clearly defined logic.

Recent research into E2E transformer models (DriveTransformer and others) has demonstrated superior performance on benchmarking tests such as CARLA and nuScenes, and there are real-time E2E vision transformer controllers developed for automotive applications.

4. Industry Applications of Deep Learning

4.1 Transportation Industry :

Deep Learning technology has shown great promise within the transportation sector through their application in the area of fully autonomous vehicles. In a broad summary article, which describes how AI, ML and DL have been implemented throughout the industry to enhance the safety, efficiency and sustainability of Transportation.

Examples of where DL has been implemented include:

Self-driving vehicles (cars, trucks, buses)

- Self-driving taxis (Waymo, Ubers robotaxi services) available within US metropolitan areas such as Dallas, Austin and Atlanta.
- Self-driving delivery robots and shuttles
- Level 2/3 Driver Assistance systems (such as Tesla's Autopilot and Mercedes' Drive Pilot found on German Autobahns)

Advanced Driver Assistance Systems (ADAS)

- Lane Keeping Assist
- Automatic Emergency Braking (AEB)
- Adaptive Cruise Control
- Blind Spot Detection
- Traffic Sign Recognition

Traffic Management / Optimisation

- Traffic Flow and Congestion prediction via models developed with DL techniques
- Support Smart Traffic Signals and provide route recommendations;

Public Transportation and Logistics (eg optimise bus schedule /faster route).

- Anticipate Demand for Transit to Reduce Empty Trips to/from Community;
- Utilise Autonomous Freight Trucks over long-distance routes on Highways).
- Overall, DL will continue to contribute to making our transportation systems safer, more efficient, and more fully automated.

4.2 Security Industry

Deep learning plays a big role in **security** systems, and some of these ideas are also used in the context of self-driving cars.

1. **Surveillance and monitoring**
 - Object and person detection in CCTV feeds using CNNs.
 - Suspicious behavior detection in train stations or airports.
 - Mask-wearing and crowd density monitoring.
2. **Biometric authentication**
 - Face recognition for access control.
 - Voice recognition.
3. **Cybersecurity for autonomous vehicles**
 - Monitoring CAN bus messages for anomalies.
 - Detecting malicious attacks on sensors or communication networks.
4. **Security of transportation infrastructure**
 - Monitoring tunnels, bridges, and highways using cameras and drones.
 - Analyzing events around self-driving cars to detect accidents and near-misses.

For self-driving cars, security is especially important because if someone hacks the AI system, it could create dangerous situations. Safety organizations and researchers emphasize making machine learning safer in high-stakes settings like self-driving cars and medical devices.

4.3 Healthcare Industry

At first, healthcare seems unrelated to self-driving cars, but there are several links:

1. **AI in medical imaging**
 - Deep learning is used to analyze X-rays, CT scans, MRIs, and more to detect diseases.
 - This is similar to how self-driving cars use CNNs to analyze camera images.
2. **Autonomous vehicles in healthcare logistics**
 - Autonomous vehicles can deliver medicines, blood samples, or medical supplies between hospitals and labs.
 - They can provide transport for patients who cannot drive themselves.
 - Articles discuss how autonomous driving may support hospital logistics and elder care.
3. **Emergency response**
 - Self-driving ambulances (in the future) could automatically drive to the nearest hospital following the fastest and safest route.
 - AI can also be used to predict where accidents are likely to happen and pre-position ambulances.

So, deep learning does not only change transportation as a standalone area but also supports healthcare and safety services through autonomous mobility.

5. Current Limitations and Challenges in Self-Driving Cars

Even though deep learning has enabled huge progress, full Level-5 self-driving (anywhere, anytime, no human attention needed) is **still unsolved**. Surveys and reports describe several major limitations.

5.1 Difficult Weather and Lighting Conditions

Deep learning perception models work best in clear daylight. They struggle when:

- It is rainy or snowy
- There is fog or heavy mist
- Sun is very low and causes glare
- It is night and only headlights are available
- The camera lens is dirty or covered

Cameras are especially sensitive to these conditions. LiDAR and radar help, but they also have limitations (for example, LiDAR can be affected by heavy rain or reflections). Deep learning models must be trained with more diverse data and better augmentation techniques to handle all possible weather conditions.

5.2 Rare Situations and Long Shot Events

There are those rare situations that occur infrequently and present a danger to the public (edge cases). Listed below are some examples:

- Someone who uses a wheelchair entering traffic.
- Traffic control officers who use hand signals to stop traffic.
- A semi-truck on its side across multiple lanes.
- Construction sites with unclear temporary signage.
- Unexpected road hazards (i.e., mattress in the middle of the road, animal in the road, homeless person sleeping on the ground).

Deep learning models rely heavily on data; they are only able to predict the outcomes of scenarios which they have previously encountered (patterns). Consequently, there are long shot (i.e., long-tail) events that are difficult to capture with adequate training data. Research conducted to determine the challenges related to autonomous vehicles identifies these events as being a major area of ongoing research.

5.3 Domain Shift and Generalization

Most models are trained on data for specific countries/cities and, when being used in other countries, they encounter a "Domain Shift" due to:

- Different Road Markings
- Different Driving Laws
- Different Types of Vehicles and Pedestrians
- Different Architectural Styles

A Model trained primarily on Routes in the US may transfer poorly to Narrow European Streets, or Indian traffic conditions for instance. Researchers have attempted to mitigate this difference through techniques such as Domain Adaption and Data Collection, however this remains an active area of research.

5.4 Regulation, Safety, and Interpretability

Self-driving vehicles are often in a life-threatening environment. Therefore, there is a risk of serious injury or death should anything go wrong. Recent reports from Several National Academy of Sciences Committee Reports as Well as Several Industry Conferences highlight the need to Improve Safety in the Machines used in High-Stakes Environments such as Self-Driving Car Systems and Medical Devices.

Missing Elements of the Regulatory Process:

- **Black Box Problem** - What Explanation do Deep Neural Nets Provide When Making Their Decisions?
- **Liability** - Who is at Fault When a Self-Driving Car Crashes - The Vehicle Manufacturer, The Software Company or The Driver?
- **Regulatory Approval** - Authority will require hard evidence that the System is Safe for Use.

The Need For Explainable AI and Safety Case Studies Exists but is Currently a Field of Research.

5.5 Data Requirements and Cost

To create dependable models, businesses must have access to the following data:

- Hundreds of thousands of km in total driving data,
- Photographs and LiDAR scans that are labelled
- Rarity annotated events

The collection and labelling of this data is very costly and time-consuming. Simulation aids in this process but does not completely eliminate the need for real-world data.

6. Future Developments and Potential Solutions

Despite these obstacles, the field of autonomous driving research and the automotive industry is still advancing rapidly. Recent studies and surveys indicate several promising avenues for continued progress.

6.1 Better Transformer-Based Perception and BEV Models

Modifying existing Transformer-based perception and BEV modelling methods will help improve Autonomous Vehicle performance. Transformers are now widely utilised for BEV representations of Autonomous Vehicles, due to their ability to:

- Merge multiple camera feeds into one integrated top-down view of the environment.
- Detect long-range interactions and relationships between objects in dense urban environments.
- Perform multiple tasks, e.g., detecting objects, performing instance segmentation, generating a map etc., within one unified model.

Improving the consistency of perception results will hopefully make using the results of the perception process more straightforward for planning-related modules.

6.2 More Advanced Sensor Fusion Techniques

More sophisticated sensor fusion methods will become more prevalent in next-generation Autonomous Vehicle systems. Autonomous Vehicles will combine multiple types of sensors into one overall multi-modal fusional model:

- Camera, LiDar, Radar, GPS, High Definition Maps, and perhaps V2X (Vehicle-to-Everything) communications.
- Transformer networks will be used as the primary means of fusing the various modalities into one overall multi-modal perception model.
- The perceived situations and weather conditions will determine when each of the sensors should be deemed trustworthy.

This will help create more reliable and adaptable perception systems, and will help mitigate failures in adverse or severe weather conditions.

6.3 Better Trajectory Prediction and Social Understanding

Modeling the future trajectory of an object goes beyond simply knowing where an object is located; it is equally important to understand the social behaviour of that object.

Examples of this include a pedestrian typically waiting at a crosswalk and a driver usually slowing down near schools. It is widely accepted that future trajectory prediction models should include social behaviours, such as how pedestrians interact with their surroundings, as well as semantics associated with the environment, and multi-agent reasoning via graph networks or transformer machines.

6.4 End-to-End Learning with Stronger Safety Controls

Growing datasets and improved hardware will likely accelerate the move toward end-to-end solutions for autonomous vehicles. There is evidence of this through research papers like Chen et al.'s paper on end-to-end autonomous driving and through newer transformer-based E2E models such as DriveTransformer that have shown strong performance in benchmarking tests.

One potential trend for future development includes:

- The use of larger scale "end to end" personality models to provide the core behaviours of an automated vehicle.
- The incorporation of explicit safety measures (i.e., shields that prevent dangerous behaviours from occurring).
- Leveraging simulation and formal verification of thousands to millions of virtual test scenarios to validate E2E models prior to real-world implementation.

6.5 Simulation, Digital Twins, and Sim-to-Real Transfer

Business create detailed simulations of the world to provide coverage in extreme (rare) situations, while protecting human life (through simulation).

There are several simulations available including CARLA, Nvidia DRIVE Sim, and others that provide conversion of various jurisdictions (cities, highways, weather, traffic).

During the development phase, researchers will use simulations to train and test their policies before using those same policies in real life (Sim-to-Real).

The creation of better simulations and new methods of applying simulations to real-world applications can dramatically decrease data processing costs and increase safety.

6.6 Safety, Explainability, and Regulation

Future technologies will most probably have:

- Ability to explain itself (the AI will be able to describe its thoughts).
- Common standards for Safety Metrics and tests to ensure the safety of autopilot systems.
- An increase in cooperative research between the various stakeholders of Autopilot Technology such as Automakers, Industry Regulators, AI Researchers, etc.

In conjunction with research, NVIDIA has developed a vision-language-action software platform (Alpamayo-R1) that allows a vehicle to explain its perception of the world using conversational English. The idea is to promote increased transparency of a vehicle's decision-making process.

7. Summary of Results

7.1 What I Did

- We have selected self-driving cars (autonomous driving) as my area of application.
- I have provided a comprehensive overview of the latest in State-of-the-Art deep learning techniques that can be utilised in the autonomous driving domain including perception, 3D detection, sensor fusion, trajectory prediction, as well as both trajectory planning and end-to-end systems.
- I also provided a high-level overview of the deployment of deep learning technologies in Transportation, Security, and Healthcare with a focus on how self-driving cars have become a major use case.
- I analysed current constraints and proposed potential future directions for research into Self-Driving Cars based on the findings presented in the most recent survey and research reports.

7.2 Key Technical Findings

The use of deep learning is the foundation for the development of Self-Driving Software today.

- Networks like CNNs and Transformers are being used in conjunction with detection, segmentation, 3D Perception, Sensor Fusion, Trajectory Prediction, and sometimes even directly related to Driving Decision Making.
- The emergence of Transformer Architecture and BEV Architecture represent a significant new trend in Autonomous Driving.
- Using Vision Transformers and BEV-based architectures such as BEVFormer provides additional ability to effectively integrate multi-camera and sensor inputs and increase performance in complex environments.
- Trajectory Prediction and Multi-Agent Reasoning are still significant challenges.
- Although deep learning models are increasingly accurate when estimating vehicle movements, it is still difficult to safely and appropriately estimate the actions of human beings on the road.

IET Research Journals

End-to-End Autonomous Driving is a promising technology, but one that has generated much controversy. End-to-end systems can streamline the entire process, allowing for high

performance, however there are many open questions surrounding the methods used to train models for implementing End-to-End Autonomous Driving.

8. Conclusion

The experience I gained while working on this project opened my eyes to how far we can take deep learning to take on the real world, in ways I'd never imagined possible. That excites me but also terrifies me — these decisions have an impact on people's lives!

Here are some of the main insights I gained from doing this project:

1. Self-driving cars are much more than one model; they comprise an entire array of numerous (potentially thousands of) different types of deep-learning developed digital objects, models, and traditional systems.
2. Collecting data, interpreting that data, and marking it correctly is just as much a challenge as developing a neural network – if not more challenging.
3. Having a model with "high accuracy" is not the only thing — there are other important factors to a successful self-driving car, including the model's safety, reliability, fairness, and compliance with laws.
4. The future of autonomous vehicles will likely consist of many elements: a mixture of end-to-end learning, powerful transformer algorithms, various forms of sensor fusion, lessons learned from many types of simulations, and rigorous external safety checks.

The most important aspect of working on this project was the shift in my perception of AI: instead of working on code for a model, I now understand that AI can be used to create solutions to problems – many of them in very unpredictable or chaotic environments.

REFERENCES :

Academic & Survey Papers (Deep Learning, Object Detection, Perception, BEV, LiDAR)

Chen, L., Papandreou, G., Schroff, F., & Adam, H. (2018). *Rethinking atrous convolution for semantic image segmentation*. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).

Chen, X., Kundu, K., Zhang, Z., Ma, H., Fidler, S., & Urtasun, R. (2015). *Monocular 3D object detection for autonomous driving*. Proceedings of the IEEE CVPR.

Geiger, A., Lenz, P., & Urtasun, R. (2012). *Are we ready for autonomous driving? The KITTI vision benchmark suite*. IEEE/CVF Conference on Computer Vision and Pattern Recognition.

Girshick, R. (2015). *Fast R-CNN*. Proceedings of the IEEE International Conference on Computer Vision (ICCV).

He, K., Zhang, X., Ren, S., & Sun, J. (2016). *Deep residual learning for image recognition*. Proceedings of the IEEE CVPR.

Hu, J., Shen, L., & Sun, G. (2018). *Squeeze-and-excitation networks*. Proceedings of the IEEE CVPR.

Qi, C. R., Su, H., Mo, K., & Guibas, L. J. (2017). *PointNet: Deep learning on point sets for 3D classification and segmentation*. Proceedings of the IEEE CVPR.

Qi, C. R., Yi, L., Su, H., & Guibas, L. J. (2017). *PointNet++: Deep hierarchical feature learning on point sets in a metric space*. Advances in Neural Information Processing Systems.

Ronneberger, O., Fischer, P., & Brox, T. (2015). *U-Net: Convolutional networks for biomedical image segmentation*. Medical Image Computing and Computer-Assisted Intervention (MICCAI).

Sun, P., Kretzschmar, H., & Sapp, B., et al. (2020). *Scalability in perception for autonomous driving: Waymo Open Dataset*. Proceedings of the IEEE CVPR.

Vaswani, A., Shazeer, N., Parmar, N., et al. (2017). *Attention is all you need*. Advances in Neural Information Processing Systems.

Xu, Y., Zhang, J., Zhou, Z., et al. (2022). *BEVFormer: Learning bird's-eye-view representation from multi-camera images via spatiotemporal transformers*. European Conference on Computer Vision (ECCV).

Zhou, Y., & Tuzel, O. (2018). *VoxelNet: End-to-end learning for point cloud based 3D object detection*. Proceedings of the IEEE CVPR.

Zou, Z., Shi, Z., Guo, Y., & Ye, J. (2019). *Object detection in 20 years: A survey*. Proceedings of the IEEE.

Object Detection (YOLO, SSD, Faster R-CNN)

Bochkovskiy, A., Wang, C. Y., & Liao, H. Y. M. (2020). *YOLOv4: Optimal speed and accuracy of object detection*. arXiv:2004.10934.

Redmon, J., & Farhadi, A. (2018). *YOLOv3: An incremental improvement*. arXiv:1804.02767.

Liu, W., Anguelov, D., Erhan, D., et al. (2016). *SSD: Single shot multibox detector*. European Conference on Computer Vision (ECCV).

Ren, S., He, K., Girshick, R., & Sun, J. (2015). *Faster R-CNN: Towards real-time object detection with region proposal networks*. Advances in Neural Information Processing Systems.

Carion, N., Massa, F., Synnaeve, G., et al. (2020). *DETR: End-to-end object detection with transformers*. European Conference on Computer Vision (ECCV).

Lane Detection, Segmentation, Scene Understanding

Pan, X., Shi, J., Luo, P., Wang, X., & Tang, X. (2018). *Spatial as deep: Spatial CNN for traffic scene understanding*. AAAI Conference on Artificial Intelligence.

Neven, D., De Brabandere, B., Georgoulis, S., & Van Gool, L. (2018). *Towards end-to-end lane detection: An instance segmentation approach*. IEEE Intelligent Vehicles Symposium.

Phillion, J., & Fidler, S. (2020). *Lift-splat-shoot: Encoding images from arbitrary camera rigs by implicitly reasoning about occlusions*. European Conference on Computer Vision (ECCV).

Trajectory Prediction & Multi-Agent Modeling

Alahi, A., Goel, K., Ramanathan, V., et al. (2016). *Social-LSTM: Human trajectory prediction in crowded spaces*. Proceedings of the IEEE CVPR.

Sun, W., Zhan, W., & Tomizuka, M. (2020). *Interaction-Aware Trajectory Prediction with Latent Landmark Representations*. Robotics and Automation Letters.

Gilles, T., Campos, V., Rupprecht, C., et al. (2021). *GNN-based trajectory forecasting for autonomous driving*. IEEE Transactions on Pattern Analysis and Machine Intelligence.

Sensor Fusion (Camera + LiDAR + Radar)

Ku, J., Mozifian, M., Lee, J., et al. (2018). *Joint 3D proposal generation and object detection from view aggregation*. IEEE/RSJ International Conference on Intelligent Robots and Systems.

Liang, Z., Yang, M., Chen, Y., et al. (2022). *BEVFusion: Multi-sensor multi-task BEV model for unified perception*. arXiv:2205.13542.

Kim, S., & Kim, D. (2020). *Deep learning-based sensor fusion approaches for autonomous driving: A review*. *Sensors*, 20(19), 5558.

Reinforcement Learning for Driving

Lillicrap, T. P., Hunt, J. J., Pritzel, A., et al. (2016). *Continuous control with deep reinforcement learning*. International Conference on Learning Representations (ICLR).

Kendall, A., Hawke, J., & Janz, D., et al. (2019). *Learning to drive in a day*. International Conference on Robotics and Automation (ICRA).

Sallab, A. E., Abdou, M., Perot, E., & Yogamani, S. (2017). *Deep reinforcement learning framework for autonomous driving*. *Electronic Imaging*.

End-to-End Driving Models

Bansal, M., Krizhevsky, A., & Ogale, A. (2019). *ChauffeurNet: Learning to drive by imitating the best and synthesizing the worst*. Robotics and Automation (ICRA).

Chen, D., et al. (2022). *Learning to drive from human demonstrations: End-to-end autonomous driving benchmarks using transformers*. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Codevilla, F., Müller, M., López, A., et al. (2018). *End-to-end driving via conditional imitation learning*. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*.

Prakash, A., Chitta, K., & Geiger, A. (2021). *Multi-modal end-to-end autonomous driving*. arXiv:2107.10205.

Sim-to-Real, Simulation, Digital Twins

Dosovitskiy, A., Ros, G., Codevilla, F., et al. (2017). *CARLA: An open urban driving simulator*. Conference on Robot Learning.

Wang, T., et al. (2020). *Learning from simulation and real world data via task-aware adversarial adaptation*. *Robotics and Automation Letters*.

Autonomous Driving & Safety, Regulation, Limitations

National Academies of Sciences, Engineering, and Medicine. (2021). *Assuring the safety of automated vehicles: Summary of a workshop*. The National Academies Press.

Kalra, N., & Paddock, S. M. (2016). *Driving to safety: How many miles of driving would it take to demonstrate autonomous vehicle reliability?* RAND Corporation.

Hendy, N., Gauterin, F., & Gauterin, M. (2024). *Challenges in autonomous vehicle safety under adverse weather conditions*. IEEE Intelligent Transportation Systems Magazine.

Industry Sources (Tesla, Waymo, Mercedes-Benz, Nvidia)

Waymo. (2023). *Waymo safety report*. Waymo LLC.

Tesla. (2024). *Autopilot & Full Self-Driving Capability*. Tesla, Inc.

Mercedes-Benz. (2023). *Drive Pilot Level-3 automated driving system*. Daimler AG.

Nvidia. (2024). *NVIDIA DRIVE: Scalable AI platform for autonomous vehicles*. NVIDIA Corporation.

Healthcare + AV Links

Mertz, L. (2020). *Autonomous vehicles in healthcare logistics*. IEEE Pulse.