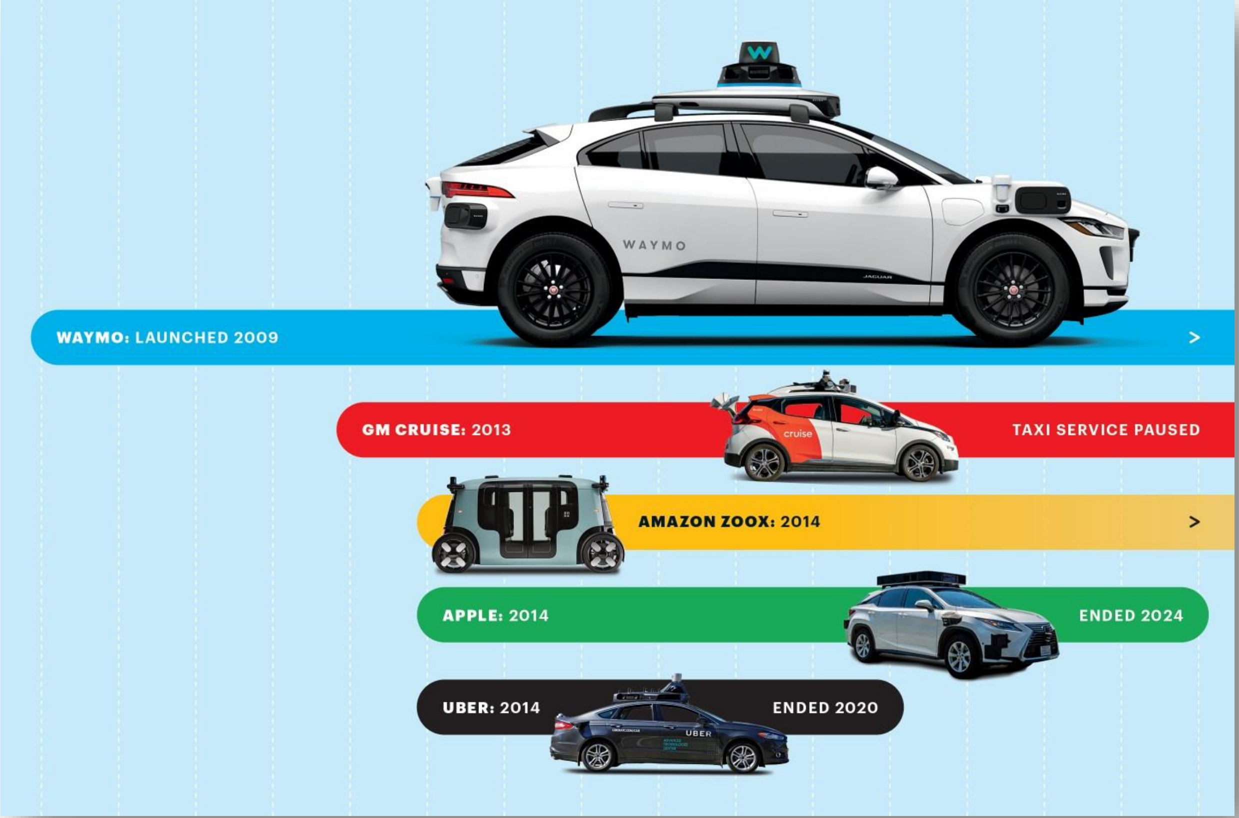
What Is AV (Autonomous Vehicle)?

An Autonomous Vehicle (AV), also known as a self-driving car, driverless car, robotaxi, robotic car, or robo-car, is a vehicle that can operate with reduced or no human input. These vehicles are designed to control their own operation, either requiring minimal input from a human driver or functioning completely without a human driver.



AVs operate using a sophisticated combination of technologies (e.g. sensing, perception, localization, planning, control).

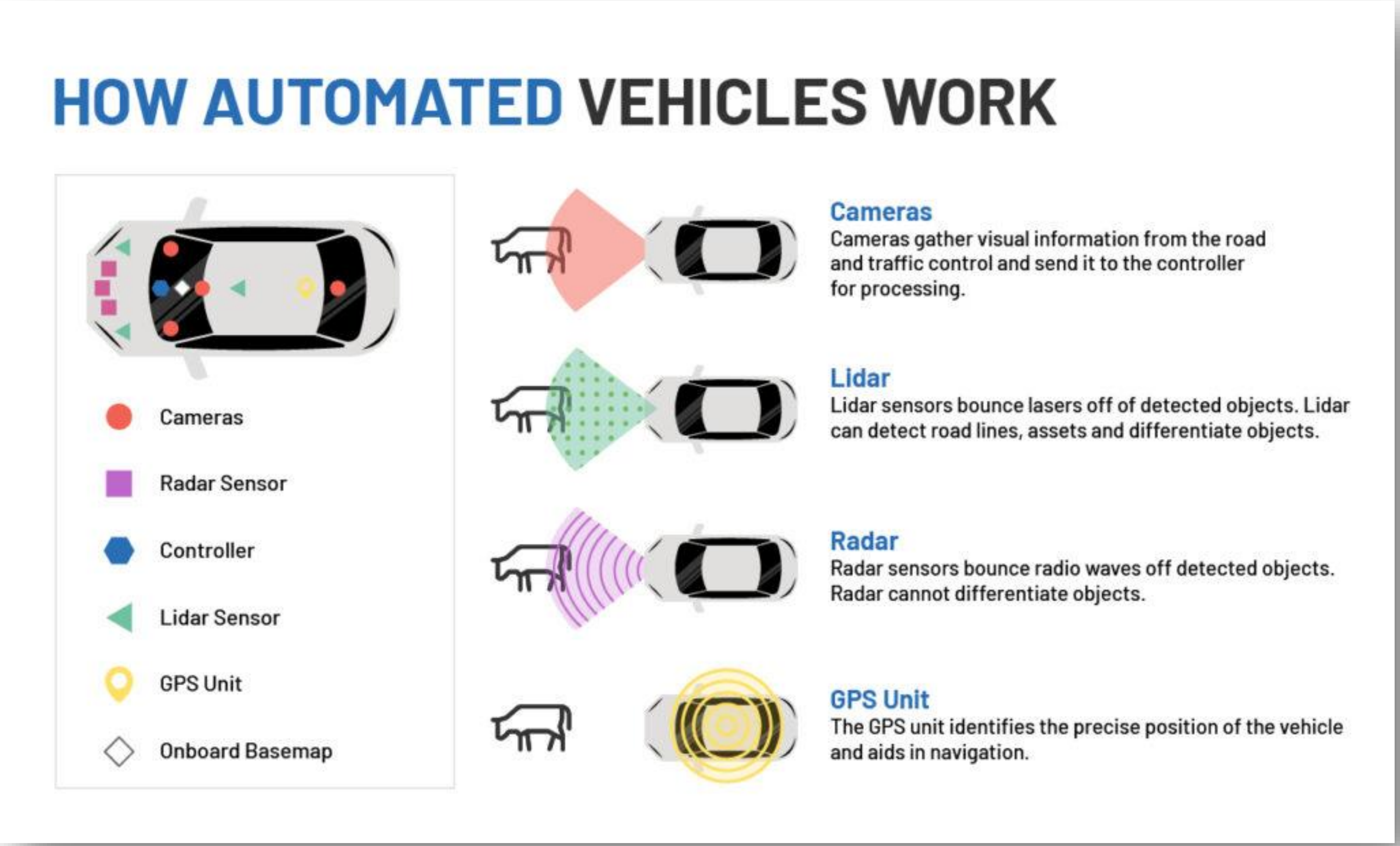
Sensing

Autonomous vehicles use various sensors that serve as the "eyes" of the vehicle:

- Cameras of various resolutions and angles, typically mounted on the windshield, bumpers, and side mirrors to provide a 360° view
- Radar
- GPS
- Lidar (Light Detection and Ranging)

These sensors monitor and create a 3-D map of the vehicle's environment, including street infrastructure, other vehicles, pedestrians, traffic lights, and road signs.

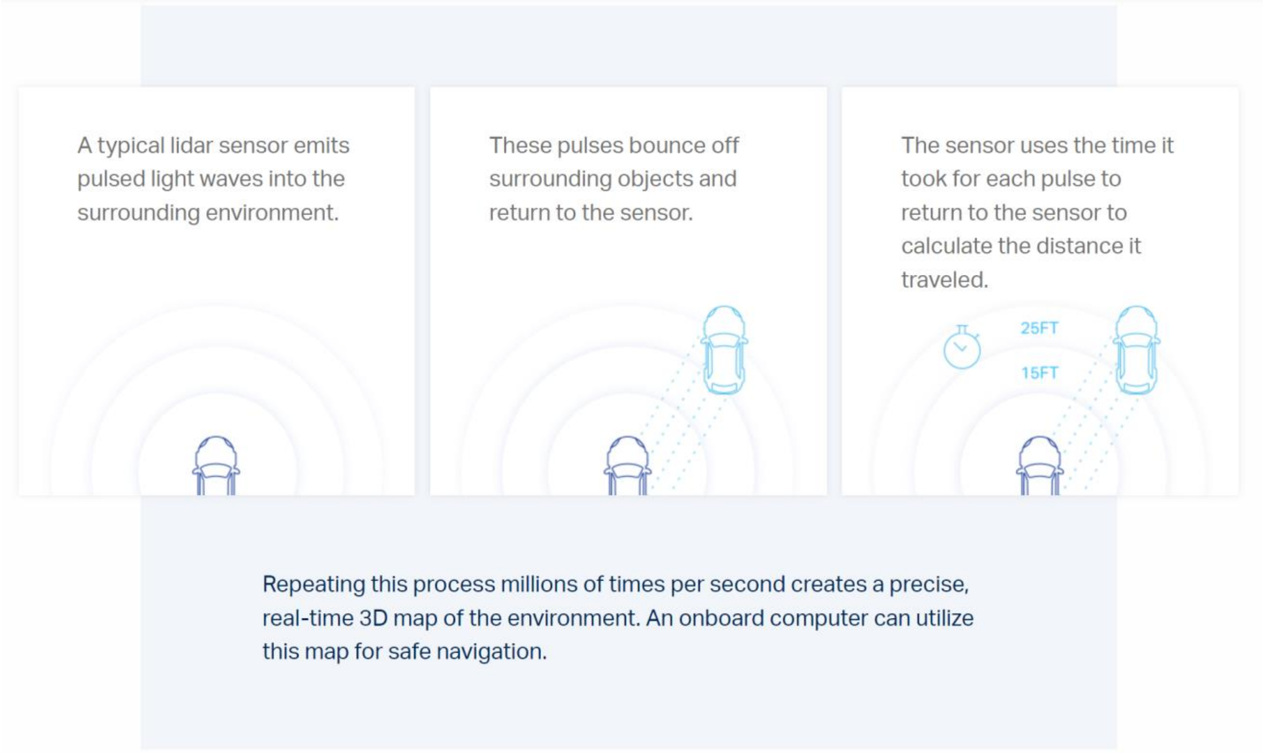
Technology Behind Autonomous Vehicles



https://zhenhub.com/blog/autonomous-vehicles/

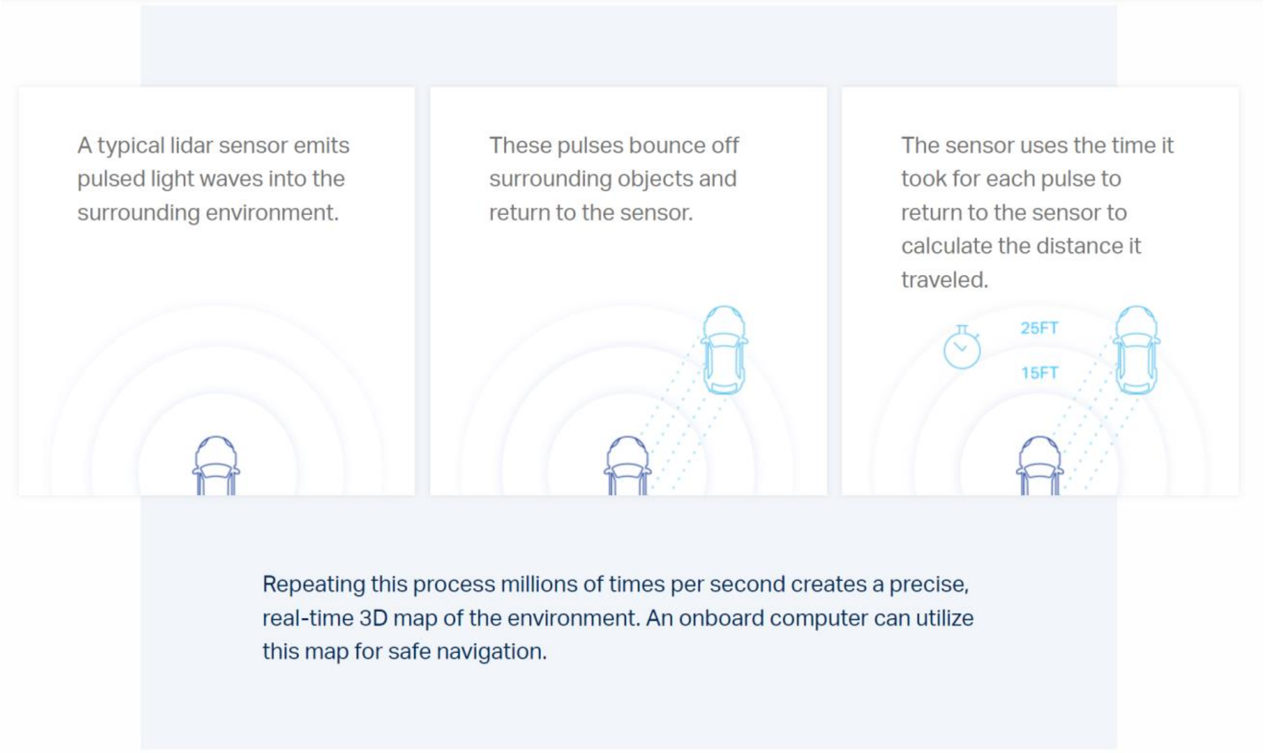
What is Lidar?

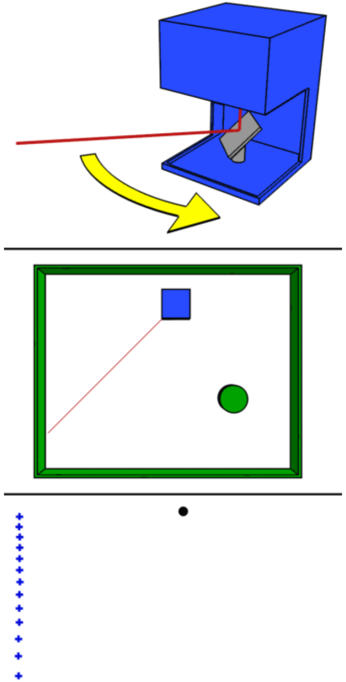
Lidar (“light detection and ranging”) uses eye-safe laser beams to “see” the world in 3D, providing machines and computers an accurate representation of the surveyed environment.



What is Lidar?

Lidar ("light detection and ranging") uses eye-safe laser beams to "see" the world in 3D, providing machines and computers an accurate representation of the surveyed environment.





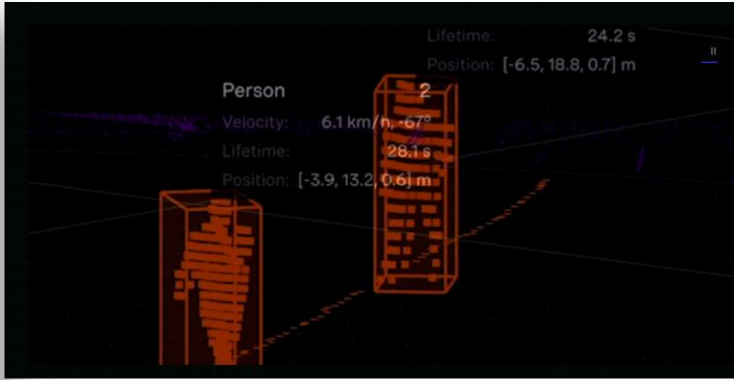
https://ouster.com/insights/what-is-lidar https://upload.wikimedia.org/wikipedia/commons/c/c0/LIDAR-scanned-SICK-LMS-animation.gif https://velodynelidar.com/products/puck/

How does lidar compare to camera systems?

Cameras produce 2D images of the environment. Lidar "sees" in 3D, a huge advantage when accuracy and precision is paramount. The laser-based technology produces real-time, high-resolution 3D maps, or point clouds, of the surroundings, demonstrating a level of distance accuracy that is unmatched by cameras, even ones with stereo vision. Whereas cameras have to make assumptions about an object's distance, lidar produces and provides exact measurements. For this reason, autonomous or highly automated systems require lidar for safe navigation. The ability to "see" in 3D cannot be underestimated. Lidar produces millions of data points at nearly the speed of light. Each point provides a precise measurement of the environment. Compared to camera systems, lidar's ability to "see" by way of precise mathematical measurements decreases the chance of feeding false information from the vision systems to the car's computer.

Camera performance is also greatly impacted by environmental conditions (e.g., bright sunlight/glare and darkness) and is therefore more susceptible to unpredictable blind spots and generating false positives or negatives. Whereas cameras are dependent on ambient light conditions and face challenges with darkness and glare, lidar provides its own light source and can therefore "see" in all lighting conditions.

Lidar has an additional technological advantage over camera systems: lidar allows the vehicle's computer to "see" the driving environment from an overhead, bird's eye perspective. The car navigates not only from a traditional driver's point of view, but can also "see" itself from the perspective of a bird flying overhead, similar to the views offered in many video games. Thus, lidar "sees" more comprehensively than a person, simultaneously looking down on all sides of the car, road, and traffic.
With accuracy and safety in mind, a lidar-centric autonomous system is a necessity. This means lidar is the central sensor and small inexpensive cameras can be added to the system for redundancy and extra care.

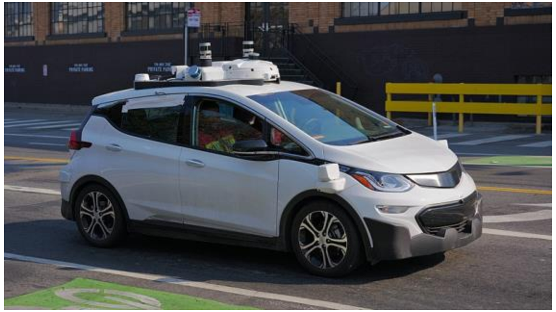


https://ouster.com/insights/what-is-lidar https://upload.wikimedia.org/wikipedia/commons/c/c0/LIDAR-scanned-SICK-LMS-animation.gif https://velodynelidar.com/products/puck/

Lidar vs Camera

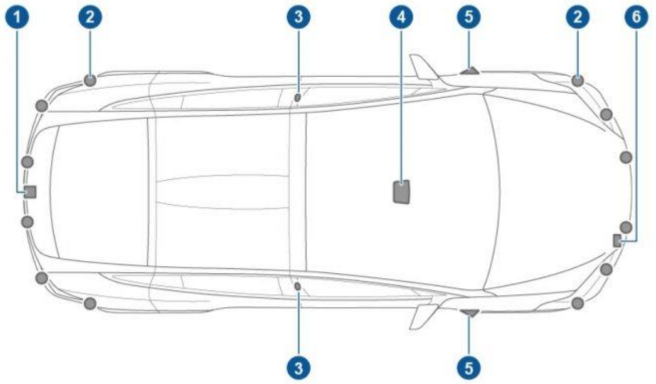
LiDAR

- cannot detect colors
- cannot interpret the text
- Impossible to identify traffic lights or road signs
- can achieve good results day and night
- high level of accuracy
- is more expensive
- requires more space
- gives self-driving cars a three-dimensional image



Camera

- can recognize colors and read road signs
- many modern AI methods to identify objects or distances
- require significantly more computing power
- camera systems are almost invisible
- challenging low-light conditions

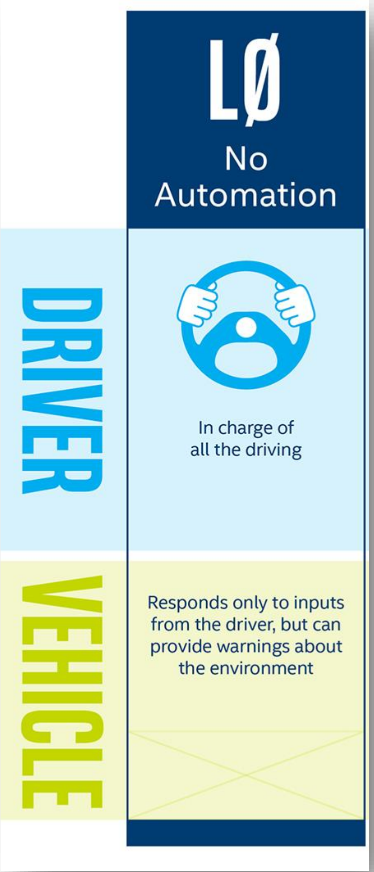


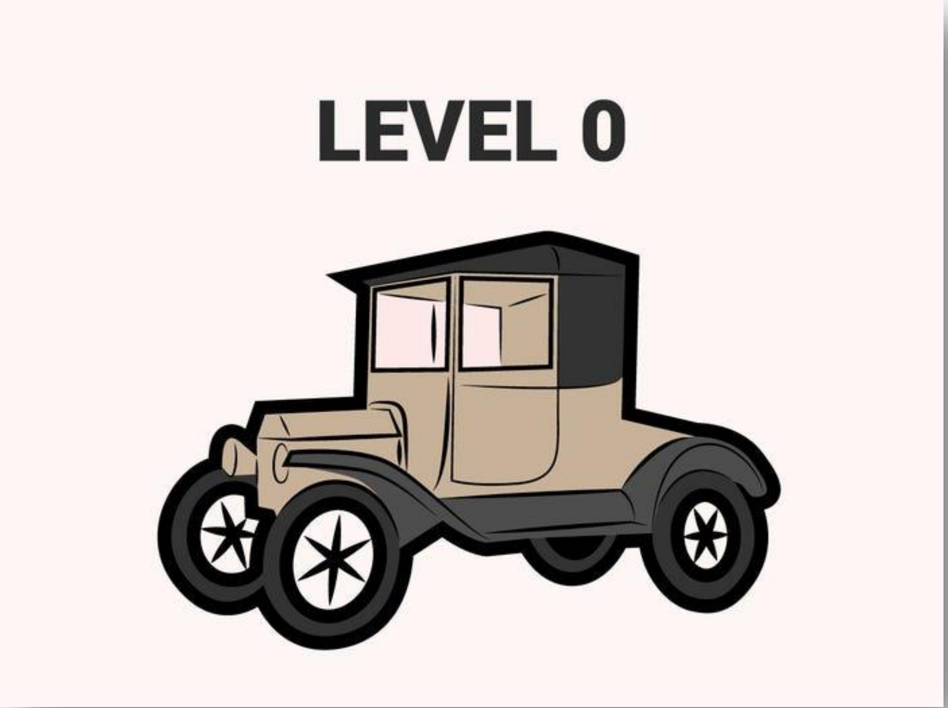
1. A camera is mounted above the rear license plate.
2. Ultrasonic sensors are located in the front and rear bumpers.
3. A camera is mounted in each door pillar.
4. Three cameras are mounted to the windshield above the rear view mirror.
5. A camera is mounted to each front fender.
6. Radar is mounted behind the front bumper on the right side of the vehicle.

Model X is also equipped with high precision electrically-assisted braking and steering systems.

https://www.autopilotreview.com/lidar-vs-cameras-self-driving-cars/
https://medium.com/0xmachina/lidar-vs-camera-which-is-the-best-for-self-driving-cars-9335b684f8d
https://leddartech.com/lidar-radar-camera-demystifying-adas-ad-technology-mix/

Levels of Autonomous Cars





Zero autonomy; the driver performs all the driving, but the vehicle can aid with blind spot detection, forward collision warnings and lane departure warnings.
https://newsroom.intel.com/news/autonomous-driving-hands-wheel-no-wheel-all

https://www.businessinsider.com/what-are-the-different-levels-of-driverless-cars-2016-10

Levels of Autonomous Cars

The vehicle may have some active driving assist features, but the driver is still in charge. Such assist features available in today's vehicles include adaptive cruise control, automatic emergency braking and lane keeping.
https://newsroom.intel.com/news/autonomous-driving-hands-wheel-no-wheel-all

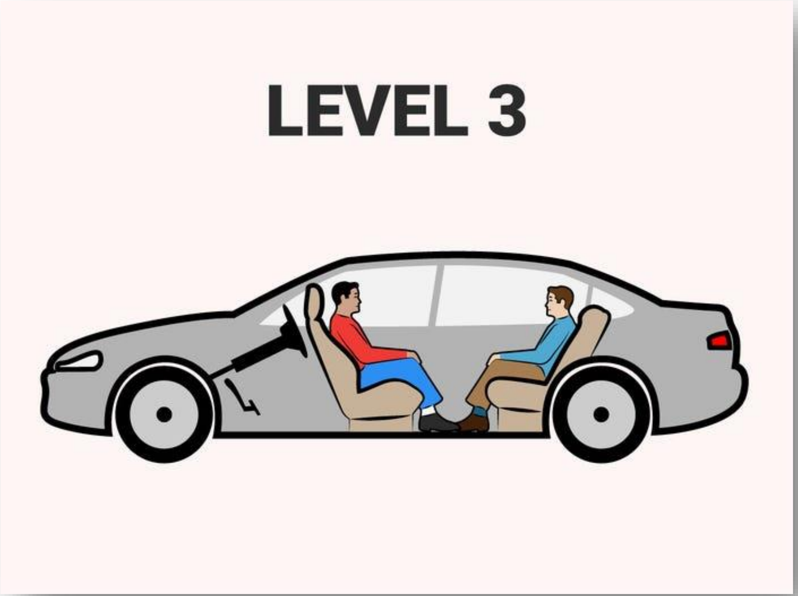
https://www.businessinsider.com/what-are-the-different-levels-of-driverless-cars-2016-10

Levels of Autonomous Cars

The driver still must be alert and monitor the environment at all times, but driving assist features that control acceleration, braking and steering may work together in unison so the driver does not need to provide any input in certain situations. Such automated functions available today include self-parking and traffic jam assist (stop-and-go traffic driving).
https://newsroom.intel.com/news/autonomous-driving-hands-wheel-no-wheel-all

https://www.businessinsider.com/what-are-the-different-levels-of-driverless-cars-2016-10

Levels of Autonomous Cars

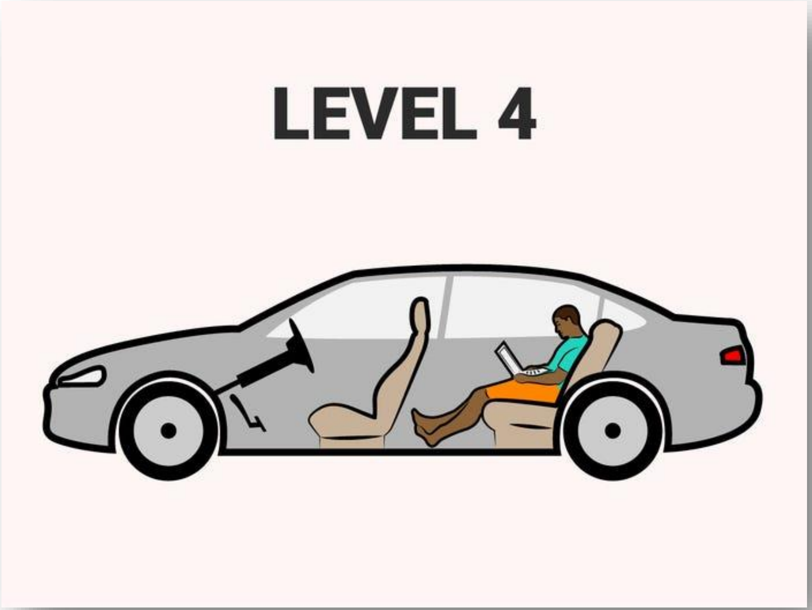


The vehicle can itself perform all aspects of the driving task under some circumstances, but the human driver must always be ready to take control at all times within a specified notice period. In all other circumstances, the human performs the driving.

https://newsroom.intel.com/news/autonomous-driving-hands-wheel-no-wheel-all

https://www.businessinsider.com/what-are-the-different-levels-of-driverless-cars-2016-10

Levels of Autonomous Cars



This is a self-driving vehicle. But it still has a driver’s seat and all the regular controls. Though the vehicle can drive and “see” all on its own, circumstances such as geographic area, road conditions or local laws might require the person in the driver’s seat to take over.
https://newsroom.intel.com/news/autonomous-driving-hands-wheel-no-wheel-all

https://www.businessinsider.com/what-are-the-different-levels-of-driverless-cars-2016-10

Levels of Autonomous Cars

https://www.businessinsider.com/what-are-the-different-levels-of-driverless-cars-2016-10

The vehicle is capable of performing all driving functions under all environmental conditions and can operate without humans inside. The human occupants are passengers and need never be involved in driving. A steering wheel is optional in this vehicle.
https://newsroom.intel.com/news/autonomous-driving-hands-wheel-no-wheel-all

Lidars vs. Cameras

Tesla CEO Elon Musk: “Anyone relying on LiDAR is doomed”

https://www.youtube.com/watch?v=HM23sjhtk4Q

Lidars vs. Cameras

Why Tesla Uses a Camera-Only Approach

Tesla's decision to rely solely on cameras stems from its belief that cameras, combined with advanced neural networks, can mimic human vision more effectively than other sensor technologies like lidar or radar. Elon Musk has argued that humans drive using their eyes and brains, so a similar setup cameras and AI should suffice for autonomous driving.

Advantages:

- **Cost-Effectiveness**: Cameras are less expensive than lidar or radar systems.
- **Scalability**: A camera-only system simplifies production and reduces hardware complexity.
- **Continuous Improvement**: Tesla leverages its global fleet to train its AI models, improving performance over time through over-the-air updates

While Tesla's camera-only approach has advantages, it also faces criticism:

- **Poor Visibility Conditions**: Cameras may struggle in rain, fog, or low-light environments where radar or lidar could perform better.
- **Depth Perception**: Critics argue that stereo or monocular vision cannot match the precision of lidar for depth estimation.
- **Edge Cases**: Situations involving visual obstructions or illusions can confuse the system, leading to potential safety risks

https://www.youtube.com/watch?v=HM23sjhtk4Q

Tesla AI Day 2021

https://www.youtube.com/watch?v=j0z4FweCy4M

[CVPR'21 WAD] Keynote - Andrej Karpathy, Tesla

https://www.youtube.com/watch?v=g6bOwQdCJrc

PyTorch at Tesla - Andrej Karpathy, Tesla

https://youtu.be/oBklltKXtDE?si=3_2E7nNKGstezrtk

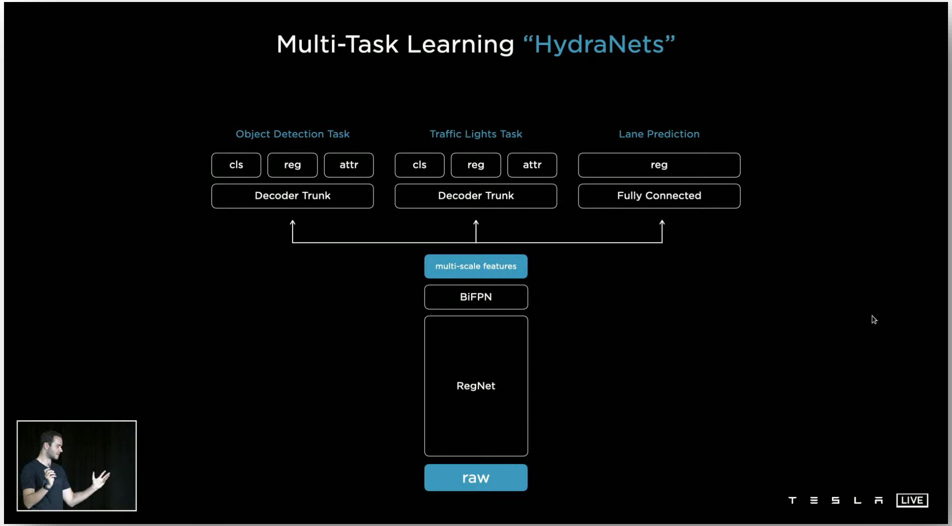
The following slides are taken from these presentations.

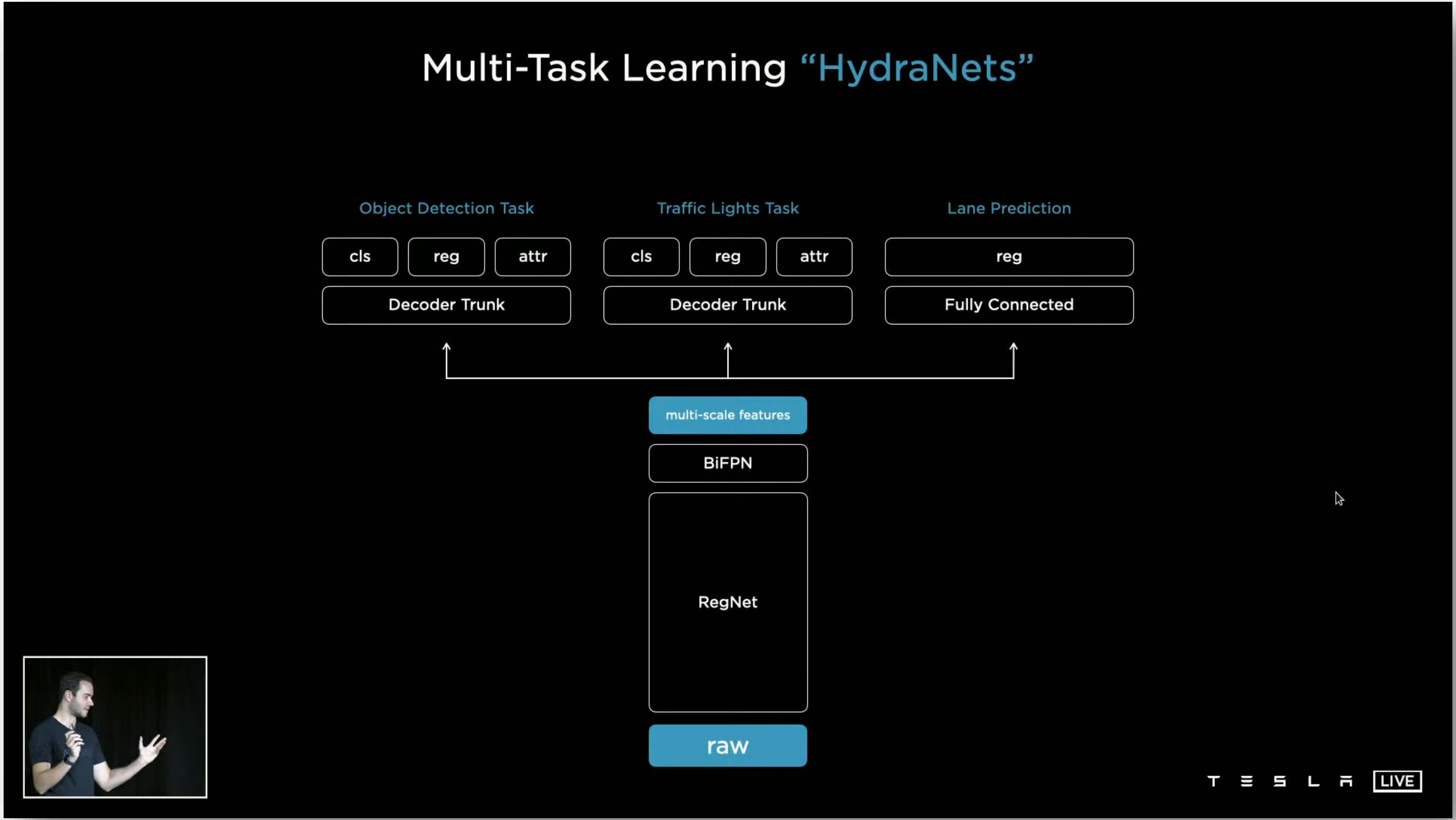
HydraNet is a neural network architecture developed by **Tesla** for their self-driving technology. This innovative approach allows Tesla to process multiple tasks simultaneously using a single, unified neural network, rather than relying on separate models for each function.

The system uses convolutional neural networks (CNNs) for image recognition and recurrent neural networks (RNNs) for temporal data processing, enabling it to make predictions based on sequences of frames.

In Tesla's implementation, it consists of:

- A shared **backbone** (the "body")
- Multiple task-specific **heads** (the "Hydra heads")

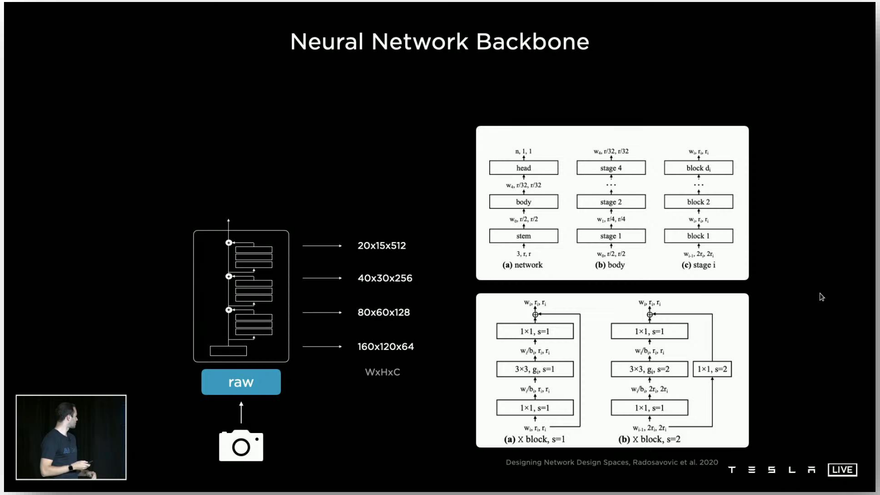


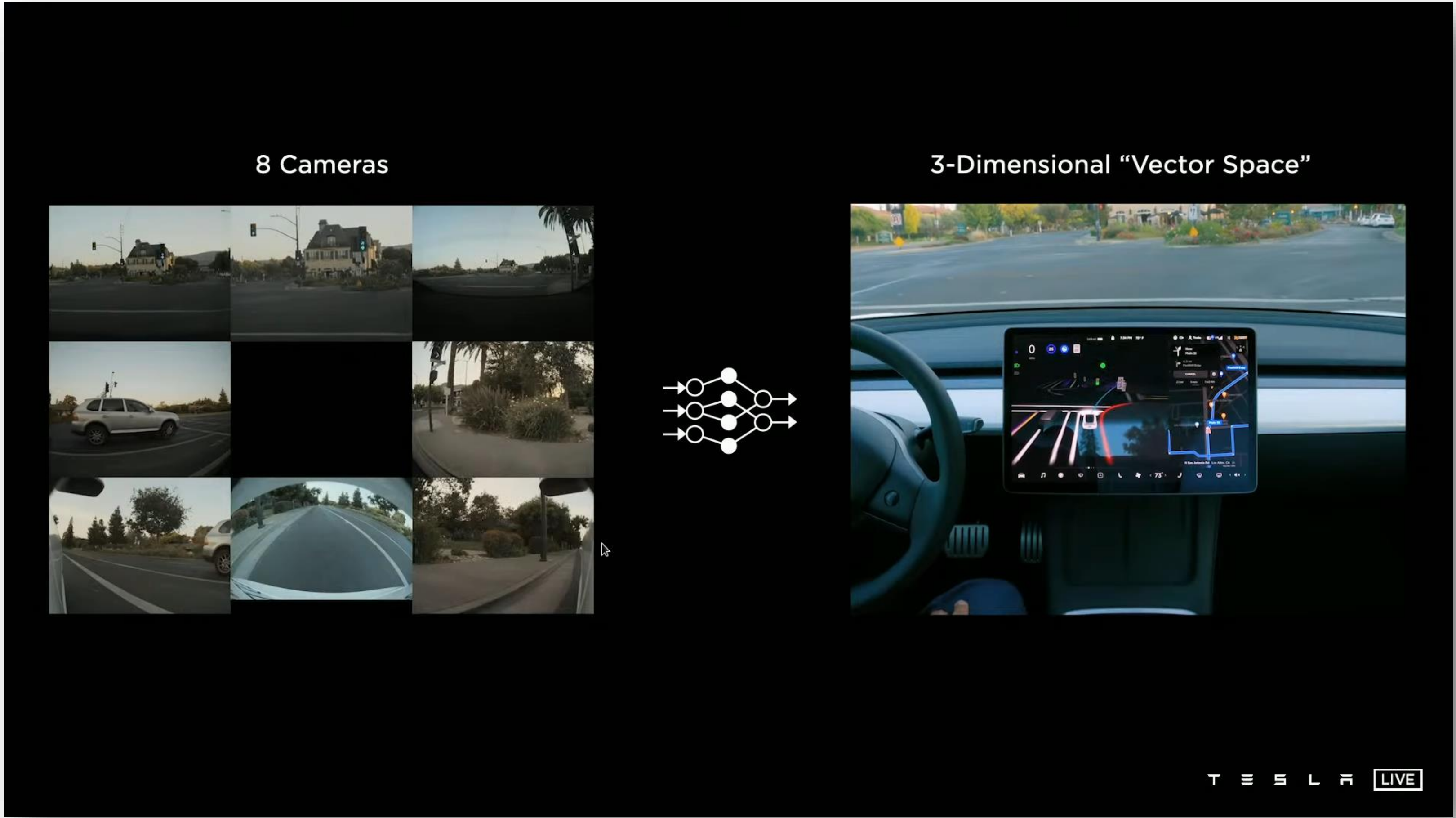

Multi-Task Learning "HydraNets"
Object Detection Task
Traffic Lights Task
Lane Prediction
cls
reg
attr
cls
reg
attr
reg
Decoder Trunk
Decoder Trunk
Fully Connected
multi-scale features
BiFPN
RegNet
raw
LIVE

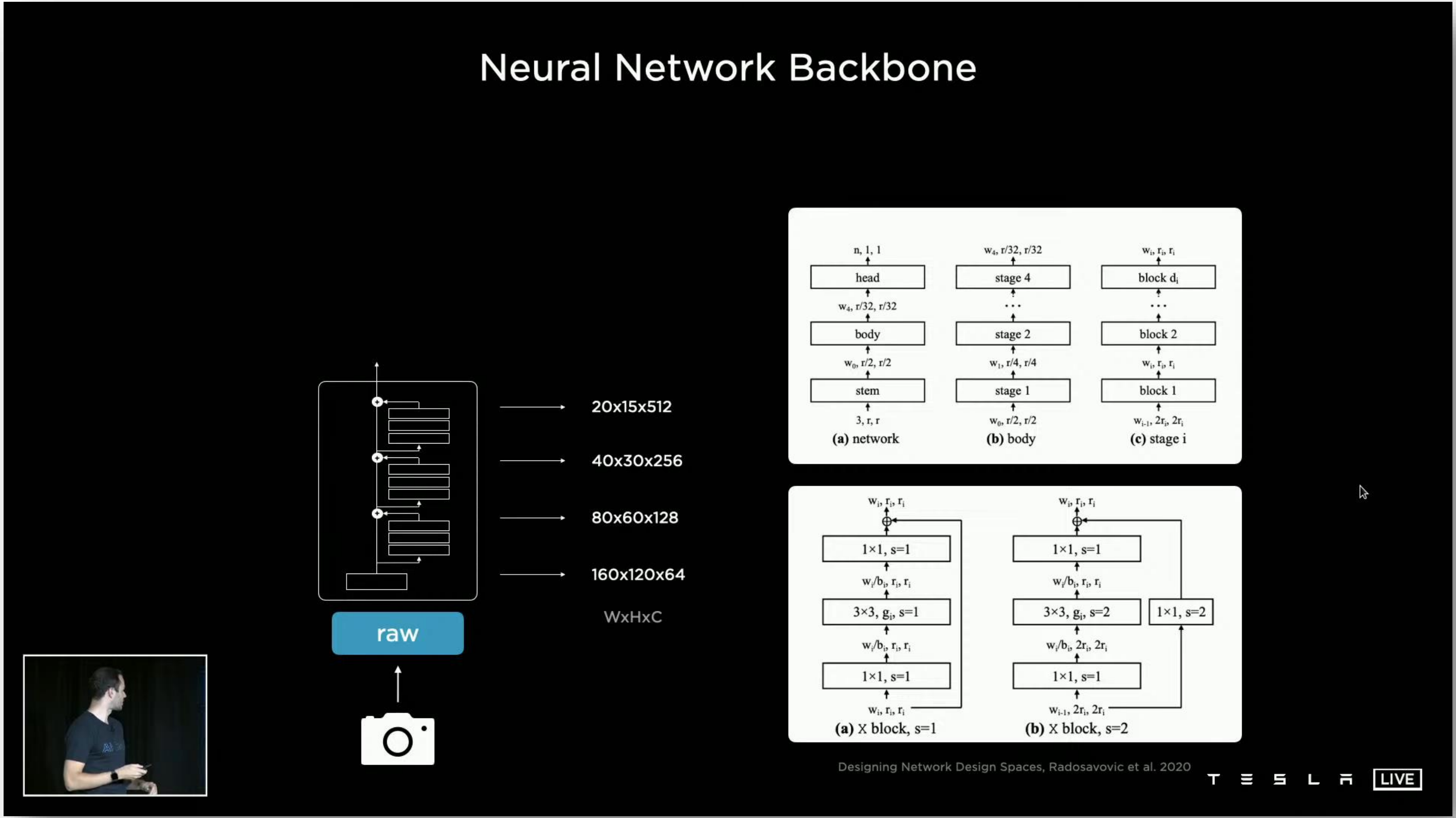
Shared Backbone

The shared backbone is responsible for processing the input data from Tesla's eight cameras. It typically includes:

1. Image Rectification: Corrects for camera calibration issues.
2. Feature Extraction: Uses a residual network (ResNet) architecture to process images into features at different scales and channels.
3. Multi-Camera Fusion: Combines data from all eight cameras into a unified representation, often using a transformer-like architecture.
4. Temporal Fusion: Incorporates time-based information by fusing current and previous frames, typically using 3D CNNs, RNNs, or Transformers




8 Cameras
3-Dimensional “Vector Space”
LIVE

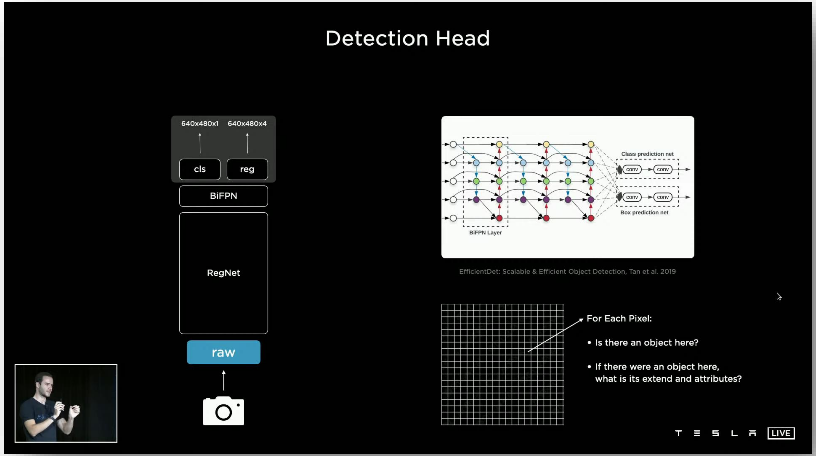

Neural Network Backbone
20x15x512
40x30x256
80x60x128
160x120x64
WxHxC
raw
n, 1, 1
head
w4, r/32, r/32
body
w0, r/2, r/2
stem
3, r, r
(a) network
stage 4
stage 2
w1, r/4, r/4
stage 1
(b) body
wi, ri, ri
block di
block 2
block 1
wi-1, 2ri, 2ri
(c) stage i
1×1, s=1
wi/bi, ri, ri
3×3, gi, s=1
(a) X block, s=1
3×3, gi, s=2
1×1, s=2
wi/bi, 2ri, 2ri
(b) X block, s=2
Designing Network Design Spaces, Radosavovic et al. 2020
LIVE

Task-Specific Heads

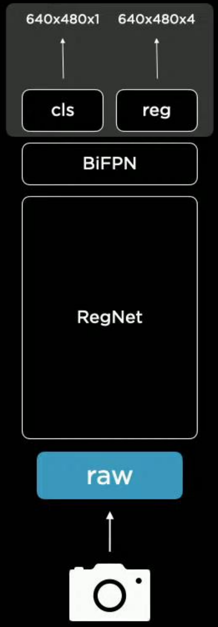
After the shared processing, the network branches out into multiple specialized heads. Each head is designed for a specific task, such as:

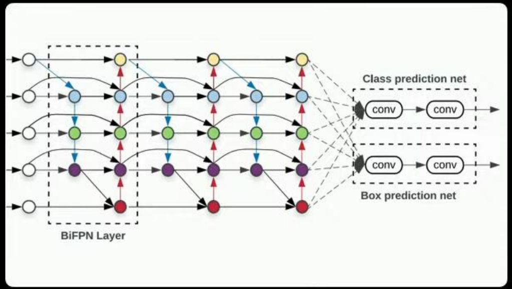
- Object detection (vehicles, pedestrians, etc.)
- Lane line detection
- Traffic light recognition
- Depth estimation
- Road layout prediction

These heads can be fine-tuned independently without affecting the performance of other tasks

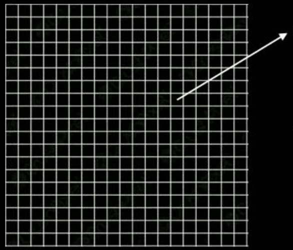


Detection Head





EfficientDet: Scalable & Efficient Object Detection, Tan et al. 2019



For Each Pixel:

- Is there an object here?
- If there were an object here, what is its extend and attributes?

TESLA LIVE

Advantages of HydraNet

1. Efficiency: By sharing a common backbone, HydraNet reduces computational overhead and allows for faster processing on Tesla's custom FSD (Full Self-Driving) computer.

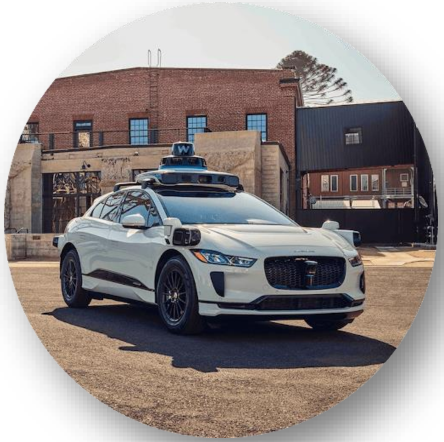
2. Flexibility: The architecture allows for easy addition or modification of task-specific heads without rebuilding the entire network.

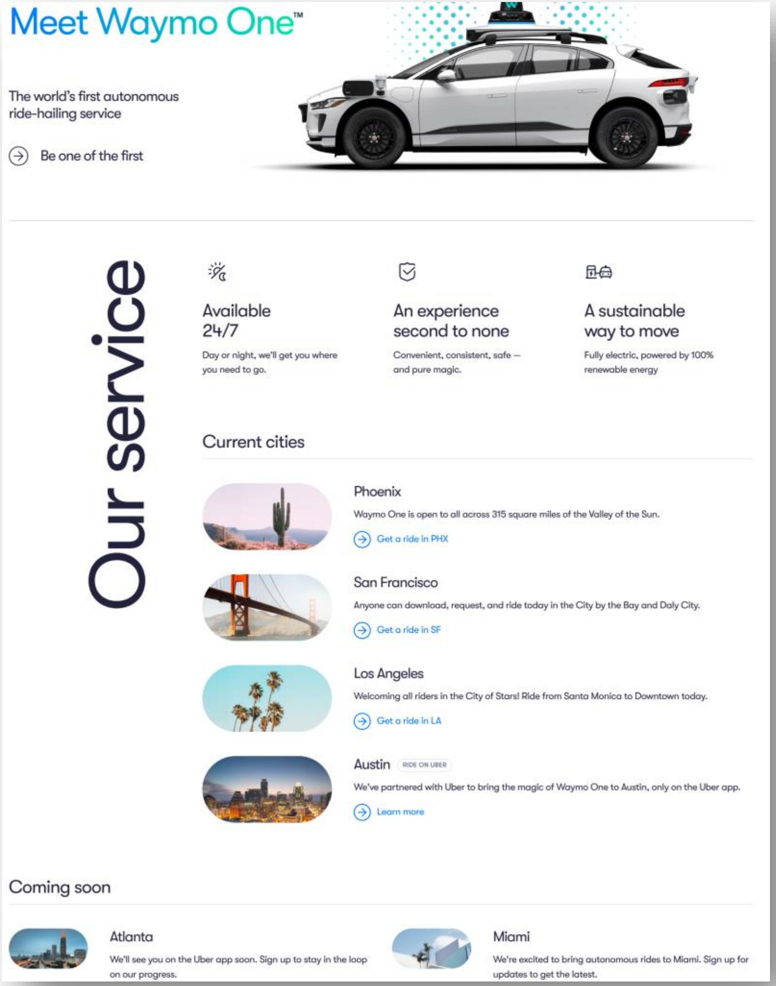
3. Multi-Task Learning: The shared backbone benefits from learning multiple related tasks, potentially improving overall performance.

4. Optimized Resource Usage: HydraNet can leverage only the necessary components for each task, e.g., lane detection might not need data from rear cameras.

Waymo https://waymo.com/

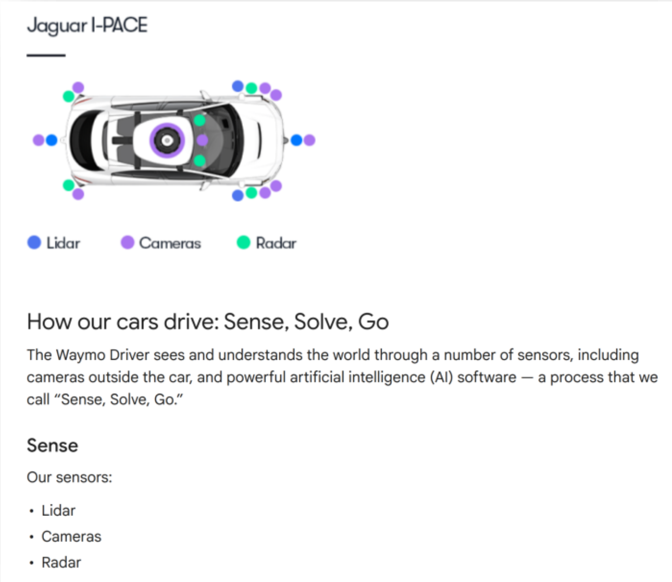
Waymo, a subsidiary of Alphabet Inc., employs a sophisticated autonomous driving system called the **Waymo Driver**. This system combines multiple types of sensors, advanced software, and machine learning algorithms to safely navigate vehicles through complex environments.





Unlike Tesla's camera-only approach (Tesla Vision), Waymo uses a multi-sensor fusion strategy that combines lidar, radar, and cameras. This approach provides higher redundancy but comes at a higher cost.

Waymo focuses heavily on detailed maps and lidar for precise localization and depth perception - a key differentiator from Tesla's reliance on neural networks trained solely on vision data. In summary, Waymo's autonomous driving system is built around a robust combination of sensors (lidar, cameras, radar), advanced AI models (prediction and planning), and extensive real-world testing.



Jaguar I-PACE

How our cars drive: Sense, Solve, Go

The Waymo Driver sees and understands the world through a number of sensors, including cameras outside the car, and powerful artificial intelligence (AI) software — a process that we call "Sense, Solve, Go."

Sense

Our sensors:

- Lidar
- Cameras
- Radar

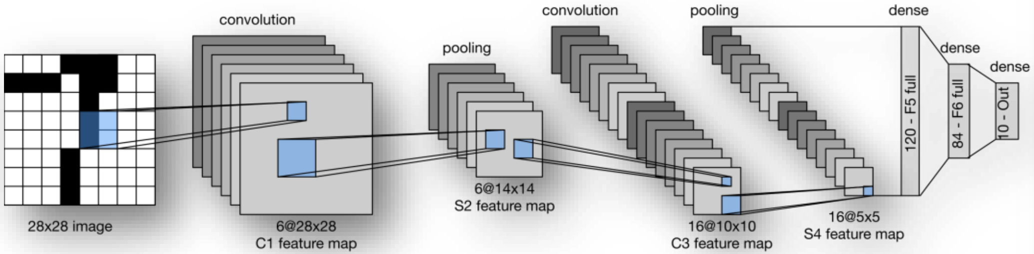
https://support.google.com/waymo/answer/9190838?hl=en

Examples of modern methods for object detection:

- R-CNN (Region based CNN, Faster R-CNN, Mask R-CNN)
  - https://pytorch.org/vision/main/models/faster_rcnn.html
- SSD – Single Shot Detector
  - https://pytorch.org/vision/main/models/ssd.html
- YOLO - You Only Look Once
  - https://github.com/ultralytics/ultralytics
- DETR (End-to-End Object Detection with Transformers)
  - https://huggingface.co/docs/transformers/main/en/model_doc/detr

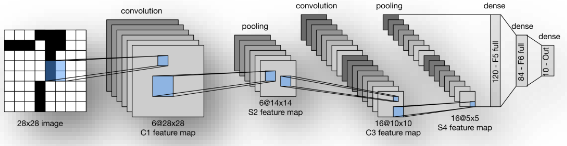
Example of CNN - LeNet:

https://pytorch.org/tutorials/beginner/blitz/cifar10_tutorial.html

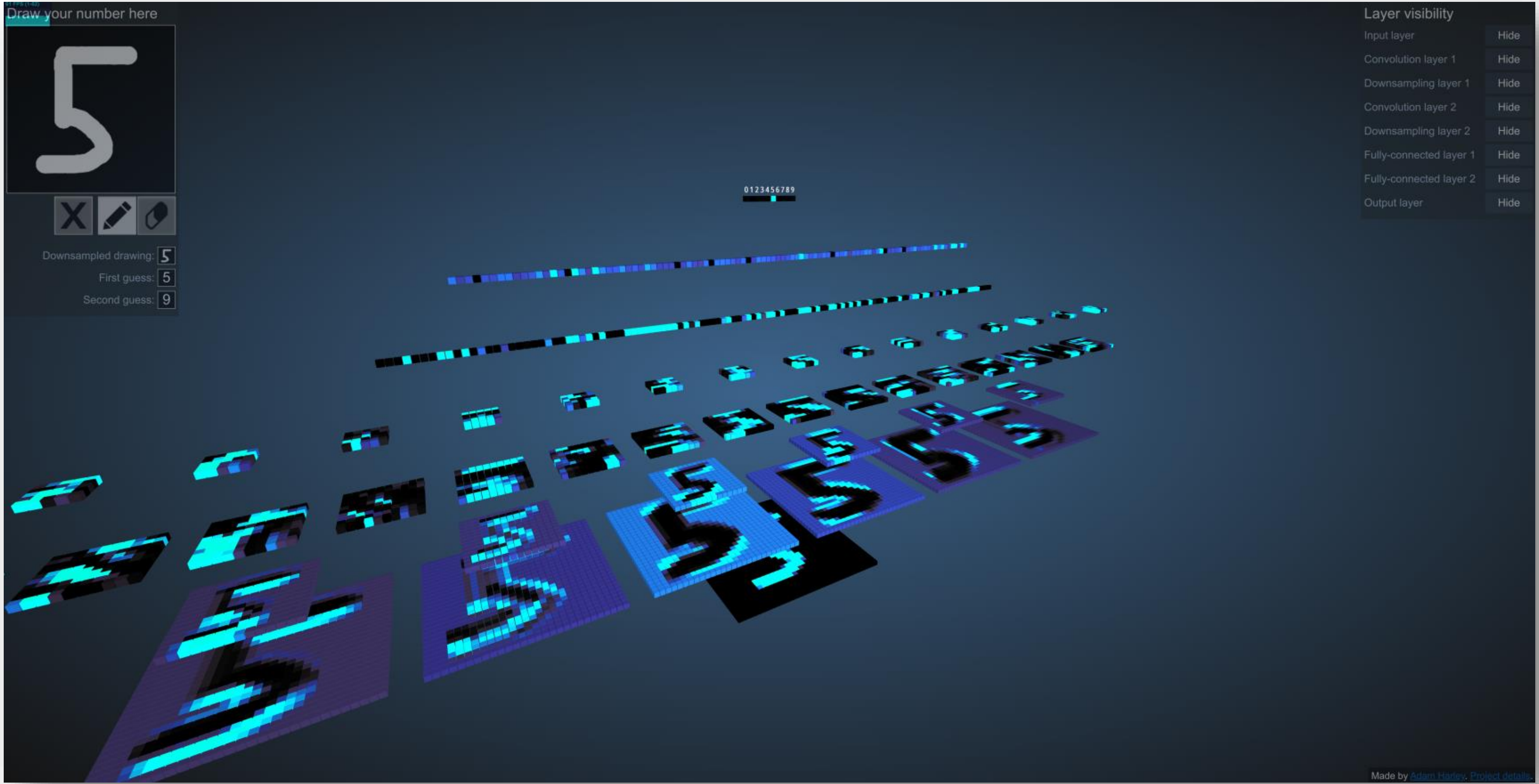



```python
import torch.nn as nn
import torch.nn.functional as F


class Net(nn.Module):
    def __init__(self):
        super().__init__()
        self.conv1 = nn.Conv2d(3, 6, 5)
        self.pool = nn.MaxPool2d(2, 2)
        self.conv2 = nn.Conv2d(6, 16, 5)
        self.fc1 = nn.Linear(16 * 5 * 5, 120)
        self.fc2 = nn.Linear(120, 84)
        self.fc3 = nn.Linear(84, 10)

    def forward(self, x):
        x = self.pool(F.relu(self.conv1(x)))
        x = self.pool(F.relu(self.conv2(x)))
        x = torch.flatten(x, 1) # flatten all dimensions except batch
        x = F.relu(self.fc1(x))
        x = F.relu(self.fc2(x))
        x = self.fc3(x)
        return x
```

https://en.wikipedia.org/wiki/LeNet



https://adamharley.com/nn_vis/cnn/3d.html



https://poloclub.github.io/cnn-explainer/