Problem Statement:

The problem statement is to find the distance of object from camera lens using single image.

- #Hash Society -

Team Members:

Amit Kumar Gupta

Akshat Apurva

Aman Kumar

Ranjit Ranjan

Problem Statement 7 :-

TRANSPORTATION IS ALWAYS BEEN A CHALLENGING IN HEAVY INDUSTRY ESPECIALLY IN MINING AND CONSTRUCTION WHERE THE ENVIRONMENT OF WORKING IS VARIABLE AND KEEP CHANGES WITH TIME. ALSO THESE INDUSTRY CATER VERY HUGE TRUCKS WHICH ALMOST A SIZE OF TWO STOREY BUILDING USED IN TRANSPORTING THE RAW MATERIAL ON WORKSITES BUT MOVEMENT OF THESE TRUCKS CREATE THE UNSAFE CONDITIONS WHICH SOMETIME LEAD TO FATALITIES IN THE INDUSTRIAL COMPLEXES. USING AI TO MITIGATE THE RISK OF PERSON COMING CLOSE TO RUNNING MACHINE OR LOGISTICS VEHICLES, DISTANCE OF OBJECT FROM MACHINE SHOULD BE KNOWN. CURRENTLY ACOUSTIC AND NON-ACOUSTIC SENSORS ARE GETTING USED TO MEASURE THE DISTANCE FROM THE MACHINE WHICH IS NOT SO INTELLIGENT TO DISTINGUISH BETWEEN THE DIFFERENT KIND OF OBJECTS. THIS NON-INTELLIGENT TECHNIQUES INCREASE THE FALSE POSITIVE ALERTS THE OPERATORS. AI COULD HELP TO CALCULATE THE DISTANCE FROM SINGLE IMAGE. ***THE PROBLEM STATEMENT IS TO FIND THE DISTANCE OF OBJECT FROM CAMERA LENS USING SINGLE IMAGE.***

Problem Solution

- We will be considering a self-Supervised approach.
- Which comprises of majorly Midas and Leres.
- Midas is a deep learning model.
- Leres is another deep learning model that can predict depth of an image.
- Both models of deep learning are used to measure distance between vehicles and people in an image of the surrounding.
- We will be converting the depth map into a real-world units using a scale factor.

Self- Supervised Approach

MIDAS
A deep learning model

LERES
A deep learning model that can predict depth

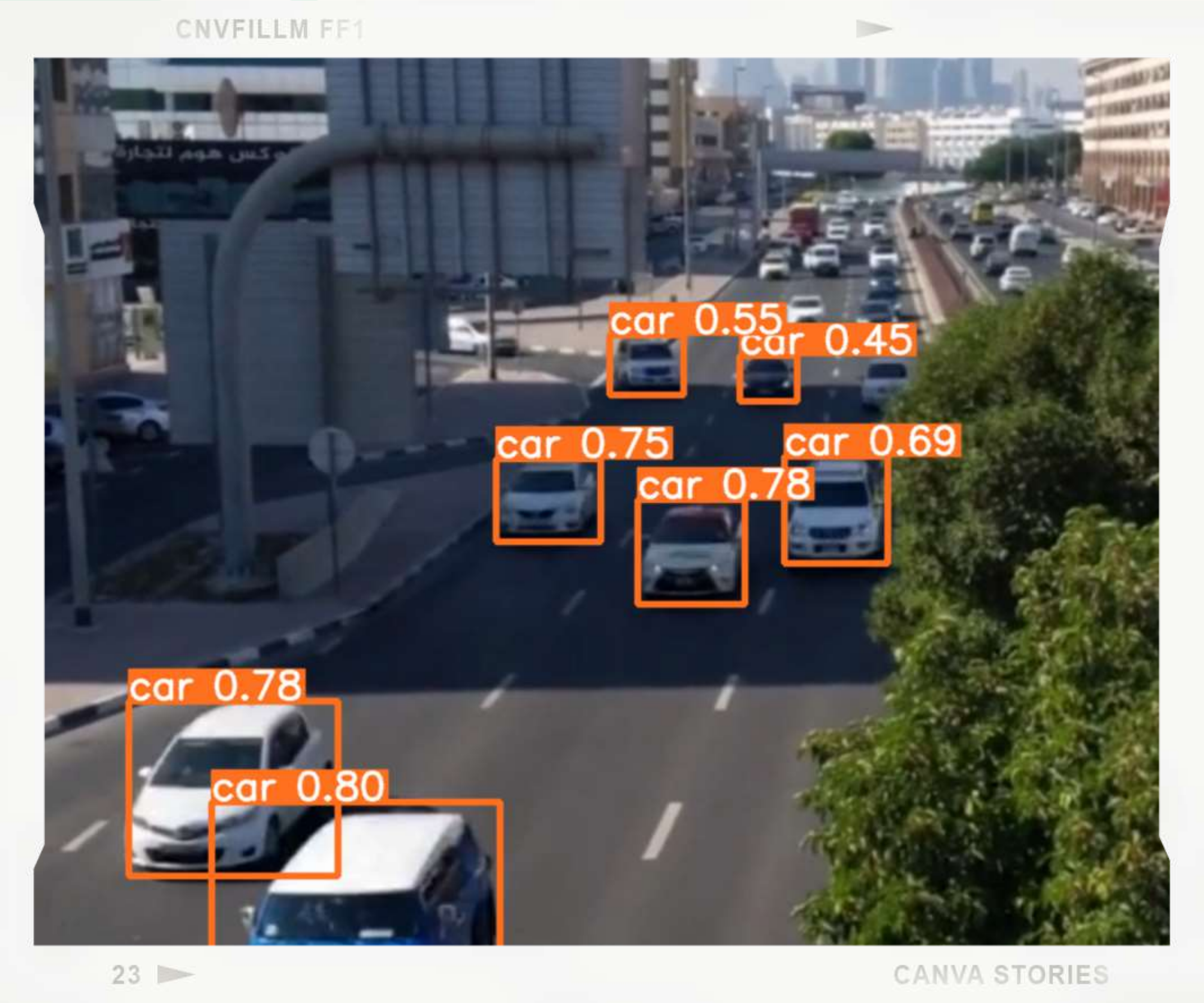
- The scale factor depends on the focal length of your camera and the resolution of your image which is being used.
- By using any tool which allows you to read pixel values from an image, such as ImageJ or Matlab.
- Accuracy of these models depends on several factors, such as the diversity and volume of the training data, the resolution of the input image, and the evaluation metric used.
- High-resolution images can be analyzed using another model that is LeRes, which can achieve a good amount of accuracy.

Our Approach includes use of :-

1. Python
2. OpenCV
3. Pytorch
4. YOLO v8n
5. KITTI dataset
6. LeRes and MiDaS model (for Depth Mapping)

We will also get much more accurate results if we feed labeled data to a Supervised Learning model. The results will get better with more input from labeled datasets such as KITTI.

YOLO implementation :-



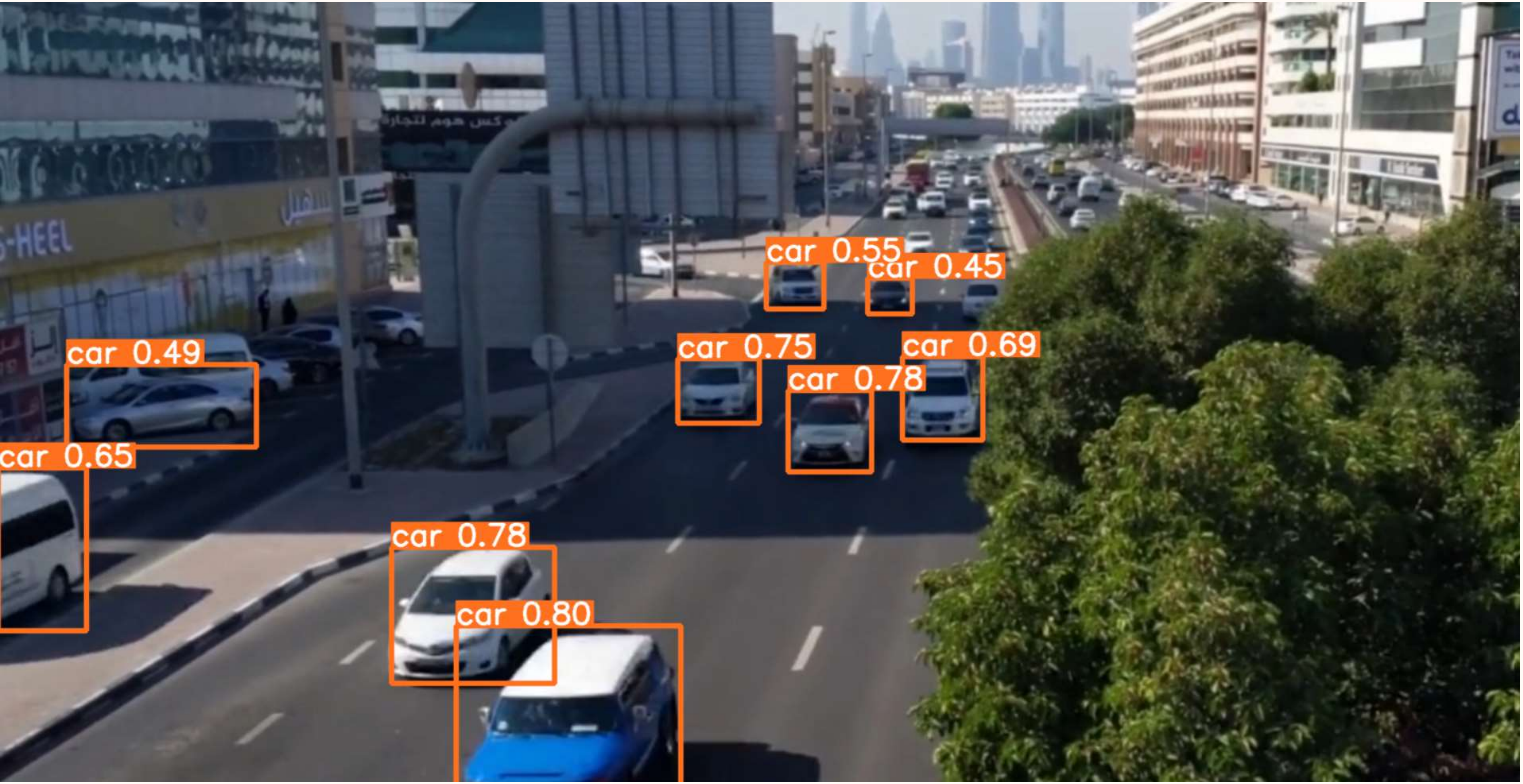
YOLO (You Only Look Once) is a real-time object detection system that uses a deep neural network to classify and localize objects in an image. Rather than dividing an image into grids and processing each grid separately, YOLO passes the entire image through the neural network at once and predicts bounding boxes and class probabilities for objects within the image. YOLO is known for its speed and efficiency, making it popular for applications that require real-time object detection, such as self-driving cars, robotics, and surveillance systems.

YOLO V8n Code :-

```
IMPORT ULTRALYTICS
IMPORT CV2
MODEL = ULTRALYTICS.YOLO('YOLOV8N.PT')

DEF COORDINATES(RESULT):
    RES = []
    FOR OBJ IN RESULT:
        BOXES  = OBJ.BOXES
        FOR BOX IN BOXES:
            RES.APPEND(BOX.XYXY[0])
    RETURN RES
WHILE TRUE:
    RESULT = MODEL('CAR_ROAD.PNG', SHOW = TRUE)
    PRINT(COORDINATES(RESULT))
```


Take it to 11
with double data.
du
S-HEEL
Al Huzaifa


car 0.55
car 0.45
car 0.49
car 0.75
car 0.69
car 0.78
car 0.65
car 0.78
car 0.80

KITTI dataset :-

- The KITTI dataset is a popular dataset for mobile robotics and autonomous driving. It contains hours of traffic scenarios recorded with various sensors, such as cameras, LiDAR, and GPS/IMU. The dataset has different tasks, such as stereo, optical flow, visual odometry, 3D object detection and 3D tracking. This dataset contains only the object detection task.
- We can use this dataset to find distance between two objects from a single monocular image. There are different methods for distance estimation from monocular images, such as using pinhole imaging, inverse perspective mapping, face and body features, or learning object-specific distance

MiDaS :-

- MiDaS is one of the most accurate models for depth estimation from a single image, especially when trained on a mixture of different datasets.
- MiDaS introduces a new loss function that absorbs these diversities, thereby eliminating compatibility issues and allowing multiple data sets to be used for training simultaneously.

LeRes :-

- LeRes model depth estimation is a project that uses a deep convolutional neural network to estimate depth from single monocular images.

- It is part of the AdelaiDepth repository on GitHub, which contains several other projects related to depth estimation.

- LeRes stands for Local Recovery Strategy, which is a technique to improve the accuracy and consistency of depth maps by using local patches and gradients.

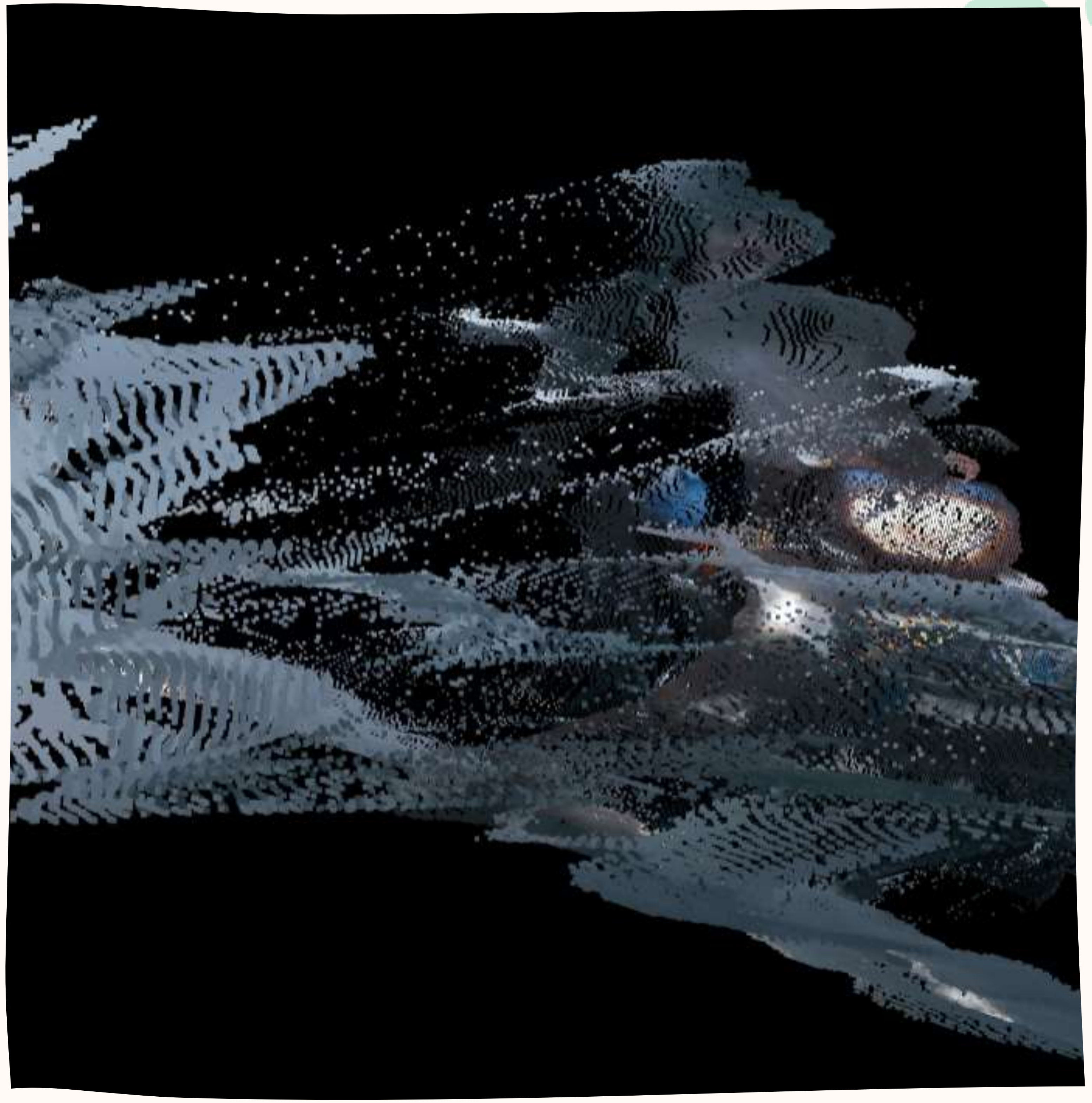
REFERENCE IMAGE

REFERENCE IMAGE

REFERENCE IMAGE

Depth Player

A depth player is a type of network that takes in both an image and its corresponding depth map as input, and produces an output that can be used for various tasks such as depth estimation, surface reconstruction, or 3D object recognition.



Video in next page :-

1. We used a web tool to visualize depth maps either as solid, mesh, or point clouds. It allows users to upload a reference image and a depth image and see the model in 3D. It is based on three.js and Kinect code. It was created by Ugo Capeto, a 3D artist and developer.
2. We can see pixels and the layers of the matrix know the distance ratio between each layer from the previous equation.

We have generated the video from https://depthplayer.ugocapeto.com/

th Player

check Depth Player blog post

Instructions

Upload a reference image and a depth image (white is near and black is far) and see the model in 3D by clicking on the "Create model" button.

Click and drag to rotate, mousewheel to zoom / Touch and drag to rotate, pinch to zoom.

Reference image

Choose File Bikes.png

Depth image

Choose File BikesD.png

Create model

Create model

Render mode

Solid Point cloud Wireframe

Actions

Download .obj

Settings

You'll have to recreate the model (by clicking on the "Create model" button) for the changes to be effective

Focal distance:	10
Near plane:	5
Far plane:	20
Smooth mesh:	0
Quad size:	1
Point size:	2
Downsampling:	2

UNDERSTANDING DISTANCE MAP :-

Our current understanding is that a depth map holds actual distance values in say meters at each pixel location while the depth image holds at each pixel location, a pixel intensity representation of that distance. So from the '**META DATA**' that we get from the photos of a camera, we are able to find

DEPTH_IMAGE = (BASELINE * FOCAL LENGTH) / DISPARITY_IMAGE) IN THE PIXEL INTENSITY DOMAIN

OR THAT DO WE MEAN INSTEAD THAT:

DEPTH_MAP = (BASELINE * FOCAL LENGTH) / DISPARITY_MAP) IN THE DISTANCE IN METERS DOMAIN.

Distance using depth map per pixels

- The only other factor you need is the height of the object in real life (otherwise you could be photographing a model which is much closer to the camera).

- The maths isn't actually that complex, the ratio of the size of the object on the sensor and the size of the object in real life is the same as the ratio between the focal length and distance to the object.

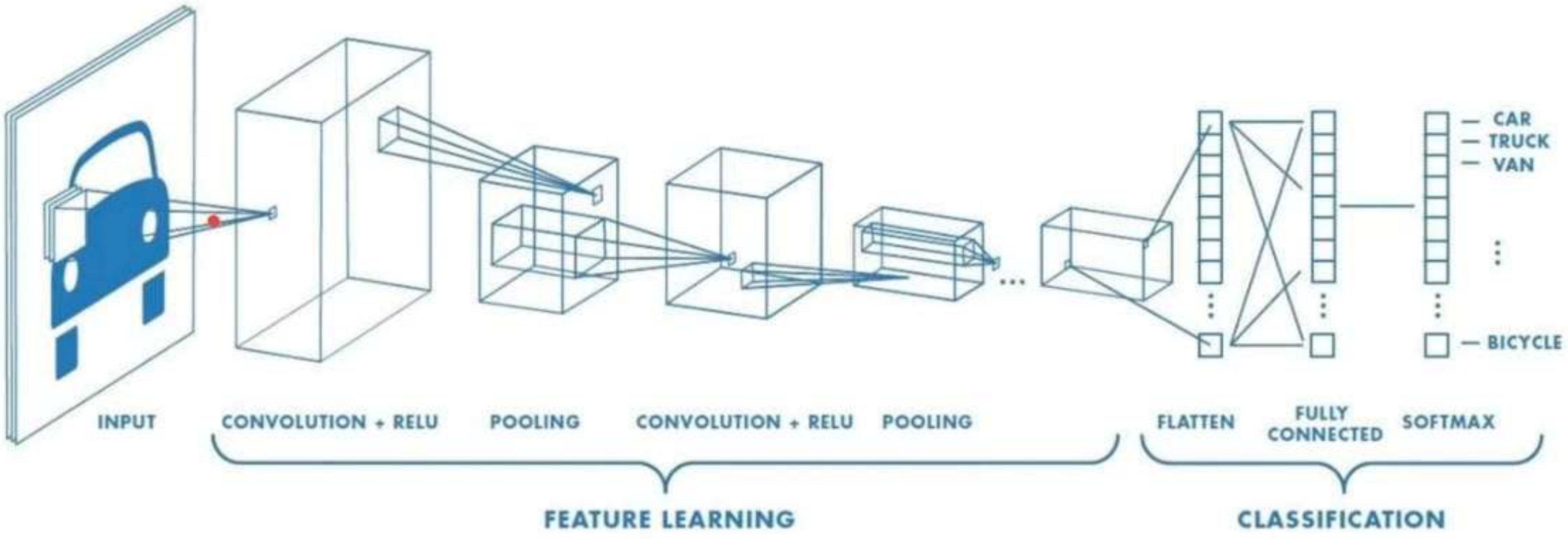
- To work out the size of the object on the sensor, work out it's height in pixels, divide by the image height in pixels and multiply by the physical height of the sensor.

REFERENCE IMAGE

Formula :-

$$\text{DISTANCE} = \frac{(\text{REAL OBJECT HEIGHT} * \text{FOCAL LENGTH})}{\text{OBJECT HEIGHT IN IMAGE}}$$

Convolutional Neural Networks



Benefits Of Yolo in Mines :-

- You can use YOLO to detect workers and vehicles in open mines. YOLO can detect different kinds of vehicles depending on predefined classes. You may need to train your own model with a custom dataset that contains images of workers and vehicles in open mines.

- You may also need to choose a suitable version of YOLO and a tracking algorithm for your task. For example, YOLO v3 and CSRT tracker have been shown to improve vehicle detection and tracking. YOLO v4 and v5 have also been developed with higher accuracy and performance.

Use of Supervised Learning Models for safety in Open Mines :-

To Detect The Angle Of Tilt Of Vehicles, You Might Need To Use A Combination Of Yolo And Supervised Learning. For Example, You Could Use Yolo To Locate The Vehicles In An Image And Then Use Supervised Learning To Estimate Their Orientation Based On Features Such As Shape, Size, Color, Or Texture. Alternatively, You Could Use Yolo To Generate Rotated Bounding Boxes (Rbbs) That Can Better Capture The Angle Of Tilt Of Vehicles Than Horizontal Bounding Boxes (Hbbs).

Steps used in Yolo Implementation :-

COLLECT AND LABEL DATA → TRAIN THE YOLO MODEL → DEPLOY THE MODEL → CONTINUOUS IMPROVEMENT

Why YOLO is preferred over R-CNN?

	YOLO v5	Faster RCNN
Inference Speed	✓	
Detection of small or far away objects	✓	
Little to no overlapping boxes	✓	
Missed Objects	✗	✗
Detection of Crowded objects	✓	✓

Finalized Procedure :-

- **Collect and label data**: Collect images and videos from the underground coal mines and label them with the objects of interest, such as people, machinery, and hazardous areas. For depth estimation, you need to have stereo images that are captured from different viewpoints. You can use labeling tools like LabelImg, CVAT, or RectLabel to label the data.
- **Train the YOLO model**: Use a deep learning framework like Darknet or PyTorch to train the YOLO model with the labeled data. You can use a pre-trained LeRes model for feature extraction in YOLO to improve its accuracy. You can fine-tune the model on your labeled data and evaluate its performance using metrics such as precision, recall, and F1 score.
- **Train the Midas model**: Use another deep learning framework like PyTorch to train the Midas model for depth estimation with the stereo images. You can use a pre-trained model and fine-tune it on your labeled data. Evaluate its performance using metrics such as mean absolute error (MAE) or root mean squared error (RMSE).

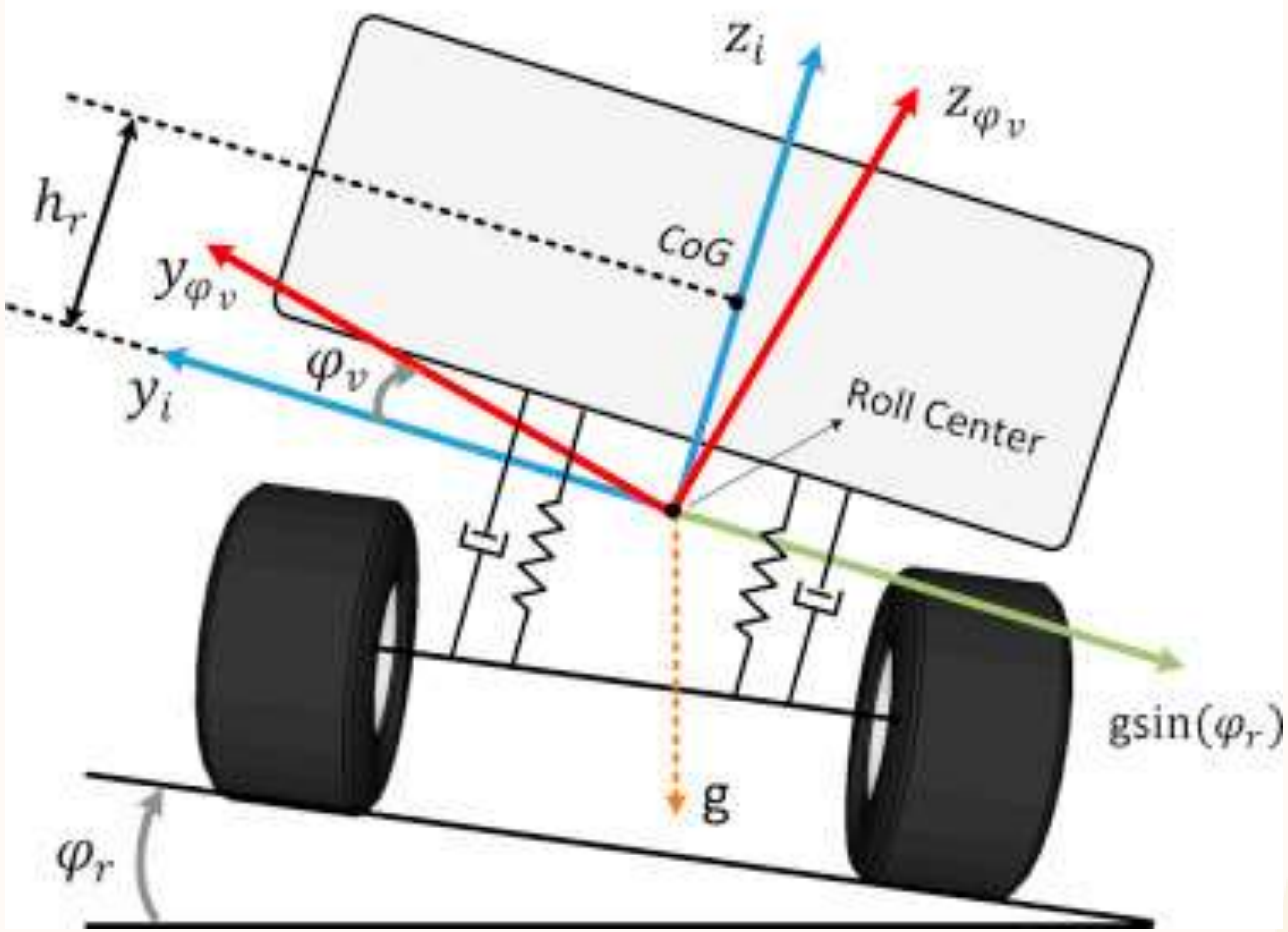
- **Deploy the models**: Once you are satisfied with the models' performance, deploy them in the underground coal mines. You can use a camera network to capture live video streams from different areas and feed them into the YOLO and Midas models for real-time object detection and depth estimation.
- **Continuous improvement**: As with any supervised machine learning model, YOLO and Midas need to be continuously updated and improved to maintain their accuracy. Collect more data from the coal mines, retrain the models, and fine-tune the hyperparameters to improve their performance.

Additional Applications :-

- The Centre Of Gravity Of A Tilting Truck Is The Point Where The Weight Of The Truck And Its Load Is Balanced. It Can Change Depending On How The Load Is Distributed And How The Truck Is Moving. If The Centre Of Gravity Moves Too Far From The Fulcrum (The Axis Of Rotation When The Truck Tips Over), The Truck Can Lose Its Lateral Stability And Overturn.

- To Prevent This, One Possible Solution Is To Use A Model That Can Detect The Angle Of Tilt Of The Truck And Adjust Its Speed, Steering, Or Suspension Accordingly. For Example, Using Yolo And Supervised Learning As Discussed Above, One Could Train A Model That Can Locate The Truck In An Image And Estimate Its Orientation Based On Features Such As Shape, Size, Color, Or Texture. Then, Using This Information, One Could Apply Some Control Logic To Keep The Centre Of Gravity Within A Safe Range.

CoG	Center of Gravity
h_r	Distance between the roll center and CoG (m)
φ_r	Road bank angle (rad)
φ_v	Vehicle roll angle (rad)



Conclusion:-

- YOLO, which stands for You Only Look Once, is a real-time object detection algorithm that uses a Deep Convolutional Neural Network to detect objects in images. It can be used for safety in mines by detecting hazards such as Explosives, Vehicles, Or Workers.

- MiDaS is a model that computes **relative inverse depth** from a single image. It stands for Multi-objective Depth Aggregation System. It can produce depth maps for different use cases ranging from high speed to high accuracy.
- LeRes is a model that estimates **depth** from a single image. It stands for Lightweight Residual Stream. It is based on ResNet-50 and uses a multi-scale feature fusion strategy. LeRes can achieve competitive performance with state-of-the-art models while being much faster and lighter. LeRes can also be used as a backbone for other tasks such as depth completion and depth super-resolution.

- Supervised Learning Is A Machine Learning Technique That Uses Labeled Data To Train A Model That Can Make Predictions Based On New Data. It Can Be Used For Safety In Mines By Classifying Images, Videos, Or Signals Into Different Categories Such As Normal Or Abnormal Conditions. Here All Are Used In Conjunction To Provide the Best Results.

Articles for improvement in YOLO for real-time calculation of distance on embedded system :-

1. Applied Sciences - Dist-YOLO: Fast Object Detection with Distance Estimation (mdpi.com)
2. Learning Depth from Single Monocular Images (neurips.cc)
3. KITTI Dataset | Papers With Code (Dataset for testing autonomous Driving)
4. Distance Map Calculation from Depth map

Thank you