REVA
UNIVERSITY
Bengaluru, India

A Project Report on

A Monocular Camera Depth Estimate Approximation using Deep learning

Submitted in Partial Fulfilment for Award of Degree of

Master of Technology
In Artificial Intelligence

Submitted By
Rajanna
R20MTA07

Under the Guidance of
Dr. Jay Bharateesh Simha
Chief Mentor, Artificial Intelligence, RACE,REVA University

REVA Academy for Corporate Excellence - RACE
REVA University
Rukmini Knowledge Park, Kattigenahalli, Yelahanka, Bengaluru - 560 064
race.reva.edu.in

August, 2022

REVA UNIVERSITY
Bengaluru, India

Candidate's Declaration

I, **Rajanna** hereby declare that I have completed the project work towards the Master of Technology in Artificial Intelligence at, REVA University on the topic entitled "A Monocular Camera Depth Estimate Approximation using Deep learning" under the supervision of **Dr. Jay Bharateesh Simha, Chief Mentor, Artificial Intelligence, RACE, REVA University.** This report embodies the original work done by me in partial fulfilment of the requirements for the award of degree for the academic 2022**.**

Place: Bengaluru	Name of the Student: Rajanna
Date: 27.08.22	Signature of Student

REVA UNIVERSITY
Bengaluru, India

Certificate

This is to Certify that the project work entitled **A Monocular Camera Depth Estimate Approximation using Deep learning** carried out by **Rajanna** with **SRN R20MTA07** is a bonafide student of REVA University, is submitting the project report in fulfilment for the award of Master of Technology in Artificial Intelligence during the academic year 2022. The Project report has been tested for plagiarism and has passed the plagiarism test with the similarity score less than 15%. The project report has been approved as it satisfies the academic requirements in respect of the project work prescribed for the said degree.

ಜೆ. ಬ. ಸಿಂಹ

Signature of the Guide	Signature of the Director
Dr. Jay Bharateesh Simha	Dr. Shinu Abhi

External Viva Panelists

Names of the Examiners

1. Dr. Santosh Nair, Founder, Analytic Edge
2. Rajkumar Dan, Data Scientist Consultant, Dell

Place: Bengaluru

Date: 27.08.22

REVA
UNIVERSITY
Bengaluru, India

Acknowledgement

I am highly indebted to **Dr. Shinu Abhi**, Director, Corporate Training for the guidance and support provides throughout the course and my project.

I would like to thank **Dr. Jay Bharateesh Simha** for the valuable guidance provided as my project guide to understand the concept and in executing this project.

It is my gratitude towards our Mentor, **Mr. Akshay Kulkarni**, and all other mentors for the valuable guidance and suggestions in learning various data science aspects and for the support. I am grateful to them for their valuable guidance on several topics related to the project.

I am thankful to my classmates for their support, suggestions, and friendly advice during the project work.

I would like to acknowledge the support provided by the founder and Hon'ble Chancellor, **Dr. P Shayma Raju**, Vice-Chancellor, **Dr. M. Dhanamjaya**, and Registrar, **Dr. N Ramesh**.
It is sincere thanks to all members of the program office of RACE who were always supportive in all requirements from the program office. It is my sincere gratitude towards my parents and my family for their kind co-operation. Their encouragement also helped me in the completion of this project

Place: Bengaluru
Date: 27.08.22

REVA UNIVERSITY
Bengaluru, India

Similarity Index Report

This is to certify that this project report titled **A Monocular Camera Depth Estimate Approximation using Deep learning** was scanned for similarity detection. Process and outcome is given below.

Software Used: Turnitin
Date of Report Generation: 25.08.2022
Similarity Index in %: 10%
Total word count: 3149
Name of the Guide: Dr. Jay Bharateesh Simha

Place: Bengaluru
Date: 27.08.2022
Verified by: M N Dincy Dechamma

Name of the Student: Rajanna
Signature of Student

Signature
Dr. Shinu Abhi,
Director, Corporate Training

List of Abbreviations

Sl. No	Abbreviation	Long Form
1	ADAS	Advanced Driver Assistance Systems
2	LDW	Generalized Linear Model
3	DHB	Dynamic High Beam
4	FCW	Forward Collision Warning
5	DL	Deep Learning
6	AI	Artificial Intelligence

List of Figures

List of Tables

Abstract

In numerous applications, such as collision detection, Robotic handling, Robotic-based manufacturing facility, and Advanced Driver Assistance Systems (ADAS), depth estimation is crucial and the most significant task. Radar, Ultrasonic, Lidar technologies both operate by reflecting radio or sound wave or laser beams respectively. Stereo cameras used for depth estimation are costly and increase the cost of the system.

There are few Monocular camera depth estimate approaches have been evolved using mathematical calculation. One approach uses pixel to depth estimate mapping and other uses geometry of road, contact point of vehicle on road and camera properties.

The proposed solution would implement depth detection using monocular camera with deep learning. The objective of the proposed solution is to detect the depth from monocular image and calibrate it with the actual depth. The pixel to distance data which derived from camera properties will be used to run through varied hidden layer and nodes to conclude the implementation using deep learning. It is evident that deeper the model that is more hidden layer and high number of nodes, faster the learning and convergence of train and test data.

Keywords: ADAS, Lidar, Radar, Ultrasonic, Deep Learning, Monocular camera, Depth estimate

Table of Contents

Chapter 1: Introduction

For depth estimate, the most prevalent methods nowadays are based on Radar [1], Lidar [2][3] or ultrasonic [4] or combination of these technologies. The radar technology measures metal object reflections and accounts for the Doppler effect in order to determine relative speed. The Lidar systems estimate distance using laser beams. Radar systems are generally relegated to high-end vehicles due to their prohibitively expensive price tags.

Monocular cameras are used in the latest driver assistance systems like Dynamic High Beam (DHB), Forward Collision Warning (FCW), Lane Departure Warning (LDW), etc. Existing systems (seen in luxury vehicles like the BMW 7 series, Mercedes E-Class, XUV700, etc.) often combine many applications into a single embedded platform to reduce hardware requirements and overall expenses. Because of this pattern, monocular camera systems are in great demand, and single-camera image processing has become obligatory.

Using stereo cameras, the conventional method of depth estimate [5] is limited in range and expensive. Some methods estimate the in-path and oblique distances of objects using a single forward-facing camera installed on the dashboard. The relationship between the 2D and actual 3D view of a image is calculated based on the camera parameters and geometry applied to the input images.

Applications that rely on a precise depth estimation include avoidance of collision and dynamic high beam assist systems. Depth estimation using a single camera is challenging because of distortions in the image caused by the camera's perspective. A significantly more precise estimate can be derived by considering the road geometry and the vehicle's contact point with the ground [6] [7][8]. The flatness of the road and the camera's optical axis being perpendicular to the road are prerequisites for this proposed method as in Figure No. 1.1.

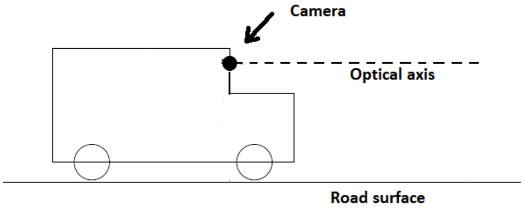


Figure No. 1.1: Schematic diagram of the camera co-planarity to road surface

1.1 Basic concept of monocular camera

Specifics of a camera include its vertical position, focal length, angular tilt, and pixel count. All other cameras are irrelevant to the ones listed here. Since the captured image is in the two-dimensional plane, distance can be determined by analyzing the similarities and differences between the flat representation and the true three-dimensional viewpoint. In order to reduce the likelihood of car crashes, the camera can be installed in a vehicle to precisely gauge the distance to an obstruction.

Since only one camera is used for data collection, perspective is used to determine distance. Size and position are two characteristics that might be used to pinpoint an object's base. Since the width of an object might fluctuate depending on its size, a distance estimate based on width will only have modest accuracy when applied to a moving vehicle, for instance. An improved estimate can be obtained by considering the road's geometry and the vehicle's contact point [6].

Under the assumption of a flat road, a camera is installed so that its optical axis is perpendicular to the road. If you place a point on the road Z metres in front of the camera, that spot will be at a certain vertical position in the final image as shown in Figure No. 1.2 [6].

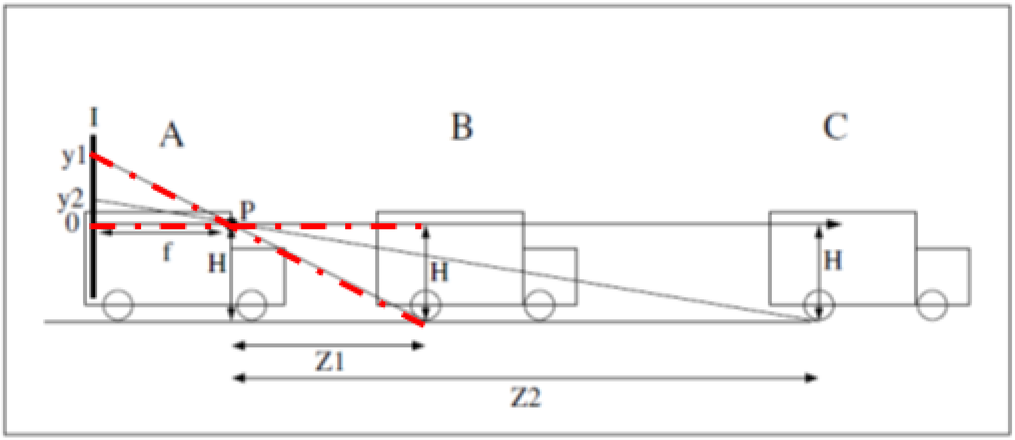


Figure No. 1.2: Schematic diagram of the imaging geometry

An image plane (I) and pinhole (P) separated by a focal distance (f) make up the schematic pinhole camera shown in Figure (2). The camera is mounted on the car (A) at a certain height (H). The back end of the car (B) is off to the side, at a distance of Z1. When viewed from above, the image plane is intersected by the vehicle's point of contact with the road (Y1). Neither the f-number nor the Y-coordinate of the image, which are typically expressed in pixels, are to scale in this illustration.

The similarity of triangles allows for a straightforward derivation of Equation (1).

$$Y/f=H/Z \qquad (1)$$

When a further away car (C) makes road contact, its contact point will be projected on its image plane at a lower (y2) location than (y1).

Chapter 2: Literature Review

Positive findings have emerged from recent studies of how artificial intelligence can be used to fields like image processing, driverless vehicles, and robotic manufacturing. One study [9][10] discusses how to take depth measurements using a single camera. Train a system to recognise a mapping between an object's pixel height and physical distance by observing and taking advantage of the correlation between these two quantities. The height of each test object in the image is then calculated by using this mapping.

Studies [6][7][11] discuss a further monocular vision system technique to determining range and measuring range. Since just one camera is used for data collection, perspective approximations are required for all distances. Both the size of the truck in the photograph and its location near the bottom of the frame could be indicators. A width-based range estimate has a 30% margin of error because the width of an unidentified vehicle (car, van, truck, etc.) could be anywhere from 1.5 m to 3 m. Using the geometry of road and the vehicle's contact point with the road, a far superior solution was provided. The assumption was a road surface that is flat and a camera with a parallel optical axis to the road surface.

In [7][12][13] for example, it is suggested that a single forward-facing camera installed on the dashboard be used to perform both in-path and oblique distance calculation.

Deep learning is a subclass of machine learning that comprises neural networks with three or more layers. These neural networks aim to imitate the behavior of the human brain, albeit their ability to "learn" from massive amounts of data falls far short of that of the human brain. Although approximations can be obtained from a neural network with a single hidden layer, the network's accuracy can be enhanced by adding additional hidden layers.

Many AI apps and services are powered by deep learning, which allows them to automate analytical and physical tasks with little to no human participation.

The technology behind deep learning is used in a wide variety of current and future products and services, including voice-enabled TV remotes, digital assistants, and credit card fraud detection (such as self-driving cars).

The first advantage of deep learning over machine learning is that feature extraction is redundant. Artificial neural networks utilizing deep learning do not require the feature extraction stage. The layers can directly and independently learn an implicit representation of the raw input as in the Figure No. 2.1.

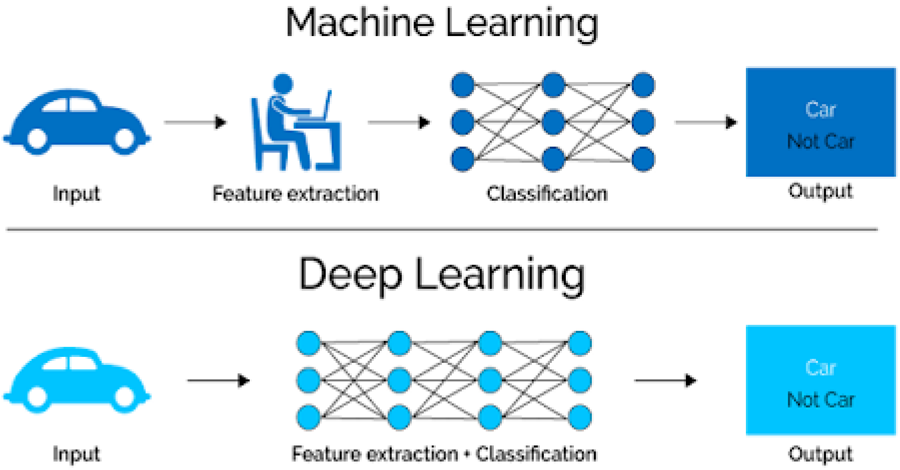


Figure No. 2.1: Machine learning Vs Deep learning (https:// builtin.com)

Various studies showed promising results in applying mathematics estimation behind the depth perspective but not many employing deep learning [14][15]. Therefore, this study addresses this research gap.

Chapter 3: Problem Statement

The Depth Estimation technique helps drivers avoid collisions using only one of the cameras often found in modern vehicles. Many different techniques have been devised to help drivers keep a safe distance between their host car and other vehicles on the road.

Some approaches make use of the correlation between the actual height of an object and its pixel size. Take use of the connection by training a system to find a mapping between the height of a pixel and the actual distance to an object. The height of each test object in the image is then calculated by using this mapping. Also, a flat road and a perpendicular visual axis to the road are taken for granted.

All of the methods use mathematical calculation to estimate the depth. None to make use of contemporary adaptations of deep learning and computer vision models.

Chapter 4: Objectives of the Study

Deep learning is a latest trend in the industry and giving breaking result in various fields may it be Text analytics, Natural language processing, computer vision and so on. DL approach in depth estimation will lead way for fast adaption of latest models and further improves the potential of industry acceptance.

The primary objective of this study is to implement the Depth estimation approximation of monocular camera system using DL model. This enables migration of legacy system to DL model which will helps better integration to AI eco system overall.

Secondary objective of the study helps in selection of optimal DL model to balance between complexity of model and cost associated with dataset of application that you are intended to use.

Chapter 5: Project Methodology

This work proposes a new way of measuring depth, one that is inspired by the method mariners use to calculate their distance from land. Navigators often guess the distance to shore based on the height of the lighthouse. Real distance is estimated by independently computing depth and lateral distance, using the same procedure as when calculating the distance between an object and its projected height (pixel height). The height of the object in pixels appears to be directly related to its perceived depth. In contrast, at any given depth, the lateral distance does not vary appreciably on the height of the item. Thus, used different technique to find a correspondence between physical and pixel-based horizontal distances as in Figure No. 5.1.

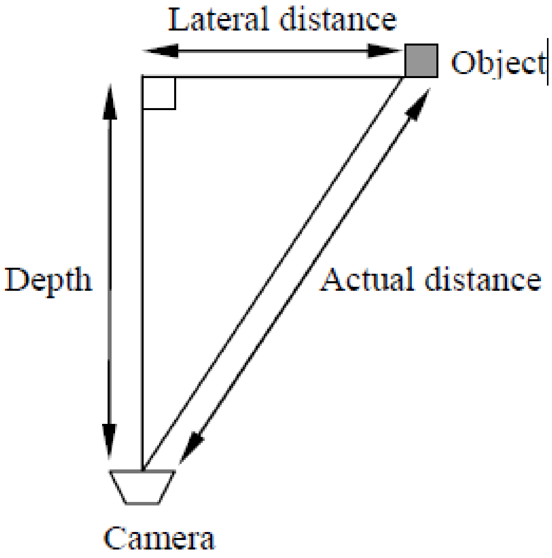


Figure No. 5.1: Top view of visualizing object distance [16].

The actual distance to an object can be determined using a Deep Learning model once both the depth and lateral distance have been determined, as shown in the following procedure.

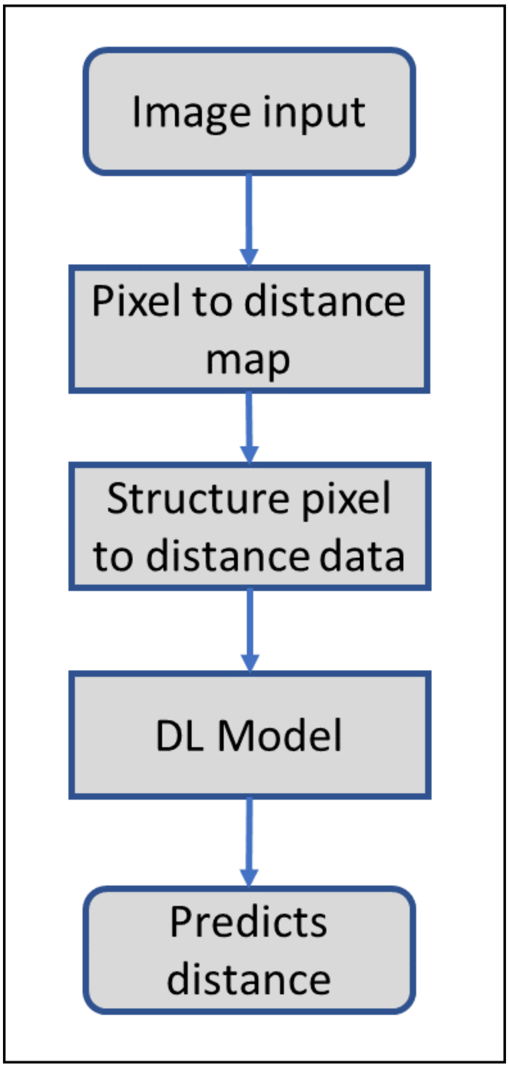


Figure No. 5.2: Proposed methodology

The workflow in the new approach is shown in Figure No. 5.2. In the new approach, DL is used to compute the depth estimation approximation.

Chapter 6: Resource Requirement Specification

There are primarily two resources that are need for the proposed solution.

1. Dataset

2. Associated python libraries for deployment of the model

6.1 Dataset

Data for the proposed implementation is simulated using camera parameters like focal length, vertical view angle, horizontal view angle and pixel density in both vertical and horizontal. Centre of the typical image as in Figure No. 6.1 used for data extraction and tabulated in Table No. 6.1.

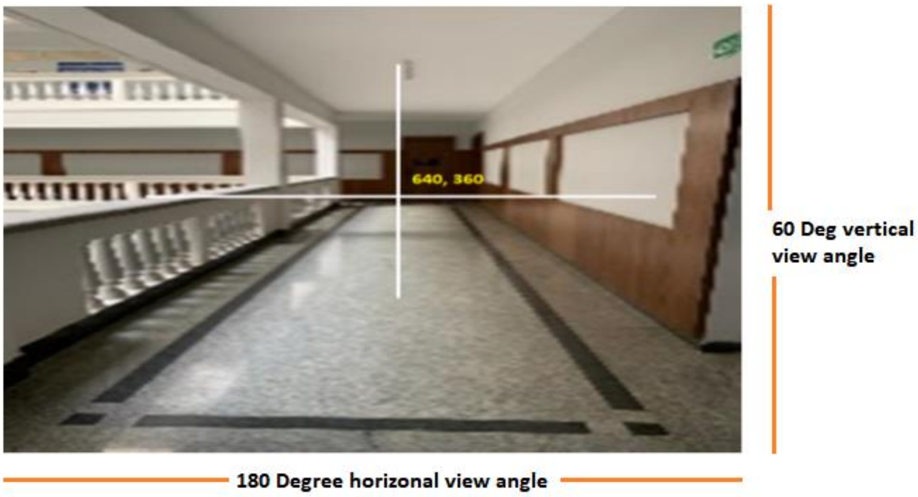


Figure No. 6.1: Pixel to distance calculation

Properties	Input variables
Input	Horizonal pixel position
	Vertical pixel position
Output	Depth estimate

Table No. 6.1: Input and Target variables for modelling

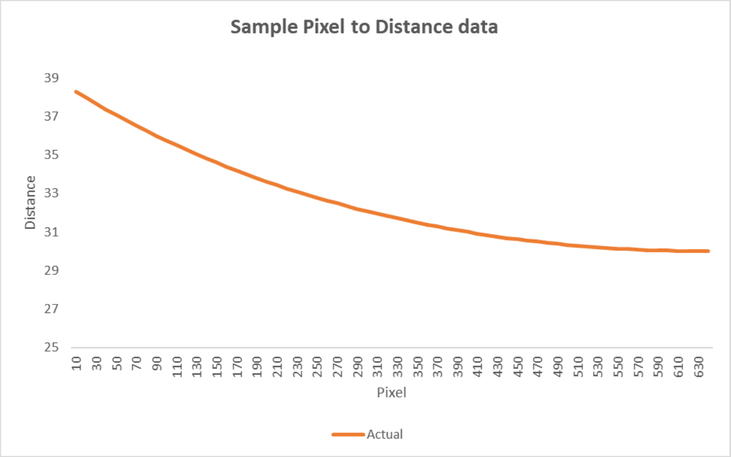


Figure No. 6.2: Sample Pixel to Distance data

It is evident from Figure No. 6.2 that relationship between distance and pixel is non-linear.

6.2 Associated python libraries for deployment of the model

The following are the requirements for this resource.

Python environment with required libraries

- Pandas for creating a data frame from dataset
- Numpy for array operation
- Sklearn for data split as train and test
- Tensorflow for loading and deploying deep learning models
- Matplotlib to run varies data representation.

Chapter 7: Implementation

Simulated data is structured as Horizonal pixel position, Vertical pixel position and Depth estimate. These are generated using camera property and pixel density. Deep learning is model is created to predict the approximate depth estimate. Implementation setup for the proposed solution is depicted in Figure No. 7.1.

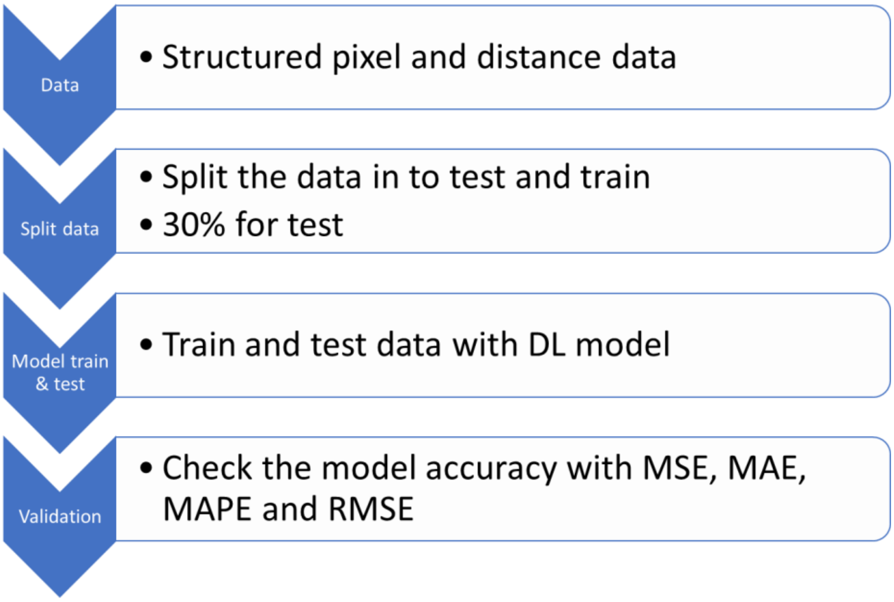

Data
• Structured pixel and distance data
Split data
• Split the data in to test and train
• 30% for test
Model train & test
• Train and test data with DL model
Validation
• Check the model accuracy with MSE, MAE, MAPE and RMSE

Figure No. 7.1: Implementation process flow

When using deep learning it is always a challenge to decide the number of hidden layer and number of nodes in each layer. Hence various trials have to be made to ensure it is optimal all respective like computation timing, accuracy, epoch..etc.

In the proposed solution, tried simple to decent complex model to ensure we arrive at optimal configuration. 4 different model configurations are tired, and compassion will be done to arrive at final solution. Note that the proposed configuration is subjected to train and test dataset.

The number of hidden layers is chosen as between 1 to 4 within each respective layer as in Table No. 7.1. The number of epochs is chosen as 200 during the model building. Rectified Linear unit (ReLu) is chosen as an activation function in all layers. Mean absolute Error (MAE) and Root Mean Squared Error (RMSE) is chosen as loss function with metric as accuracy. Adam algorithm is chosen as an optimizer.

	Node at Layer 1	**No. of hidden layers**	**No. of nodes in hidden layer**
Model 1	13	1	10
Model 2	13	4	10
Model 3	50	1	100
Model 4	50	4	100

Table No. 7.1: Model configurations for test and validation

Chapter 8: Testing and validation

As discussed in the previous chapter, the data is split into train and test using sklearn “train_test_split” method into the ratio of 70:30.

Train and test data are run through various DL model with having varied hidden layer and node for comparison of the model as mentioned in the table 7.1.

The training data used to build the DL model and validated against test data. DL model performance is validated using some of the statistical metrics such as Mean absolute Error (MAE)

- Root Mean Squared Error (RMSE)

Mean Absolute Error (MAE) is the average of the difference between the Original Values and the Predicted Values. It measures the deviation between the predicted and actual outcomes. However, they provide no information regarding the direction of the inaccuracy, i.e. whether we are under or over predicting the data. It is mathematically represented as in Equation (2):

$$MAE = \frac{1}{N}\sum_{j=1}^{N} |y_j - \hat{y}_j| \qquad (2)$$

The standard deviation of the residuals is the Root Mean Square Error (RMSE) (prediction errors). Residuals are a measure of the deviation of the data points from the regression line, and RMSE is a measure of the dispersion of the residuals. To rephrase, it tells us how closely the data fits along the line of best fit. It is mathematically represented as in Equation (3):

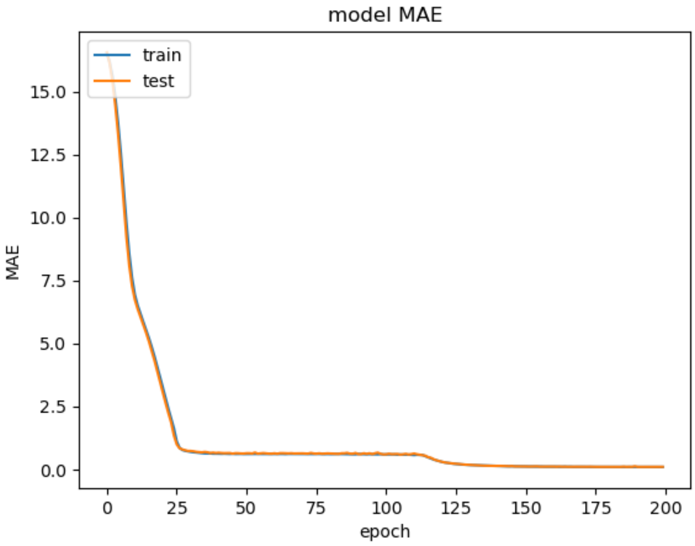
$$RMSE = \sqrt{\sum_{i=1}^{N} \frac{(y_j - \hat{y}_j)^2}{N}} \qquad (3)$$

The training and testing accuracy metrics of all the four DL model discussed are mentioned in the below Figures No 8.1 through 8.4 and Table No. 8.1:

	MAE		**RMSE**	
	Train	Test	Train	Test
Model 1	2.0245	1.8609	2.6184	2.4958
Model 2	0.5060	0.5226	0.6550	0.6694
Model 3	0.1605	0.1675	0.2327	0.2494
Model 4	0.0623	0.0630	0.0837	0.0839

Table No. 8.1: Train and test metrics for varies model configurations at Epoch 25

Model 1 Train and test metrics:



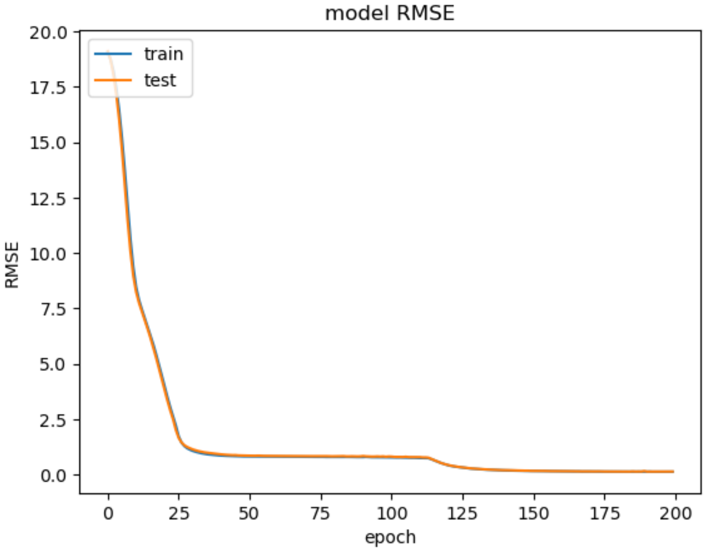


Figure No. 8.1: Model 1 Train and Test metrics

Model 2 Train and test metrics:

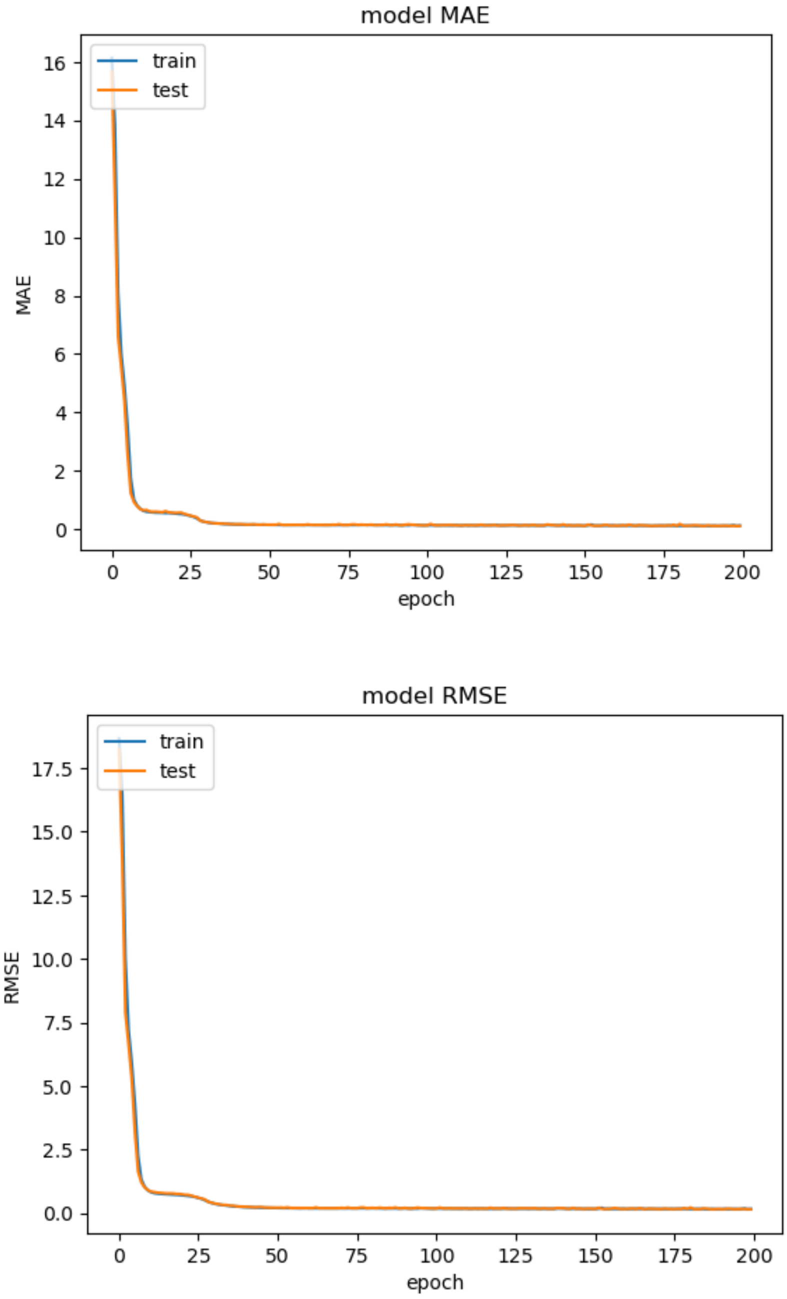


Figure No. 8.2: Model 2 Train and Test metrics

Model 3 Train and test metrics:

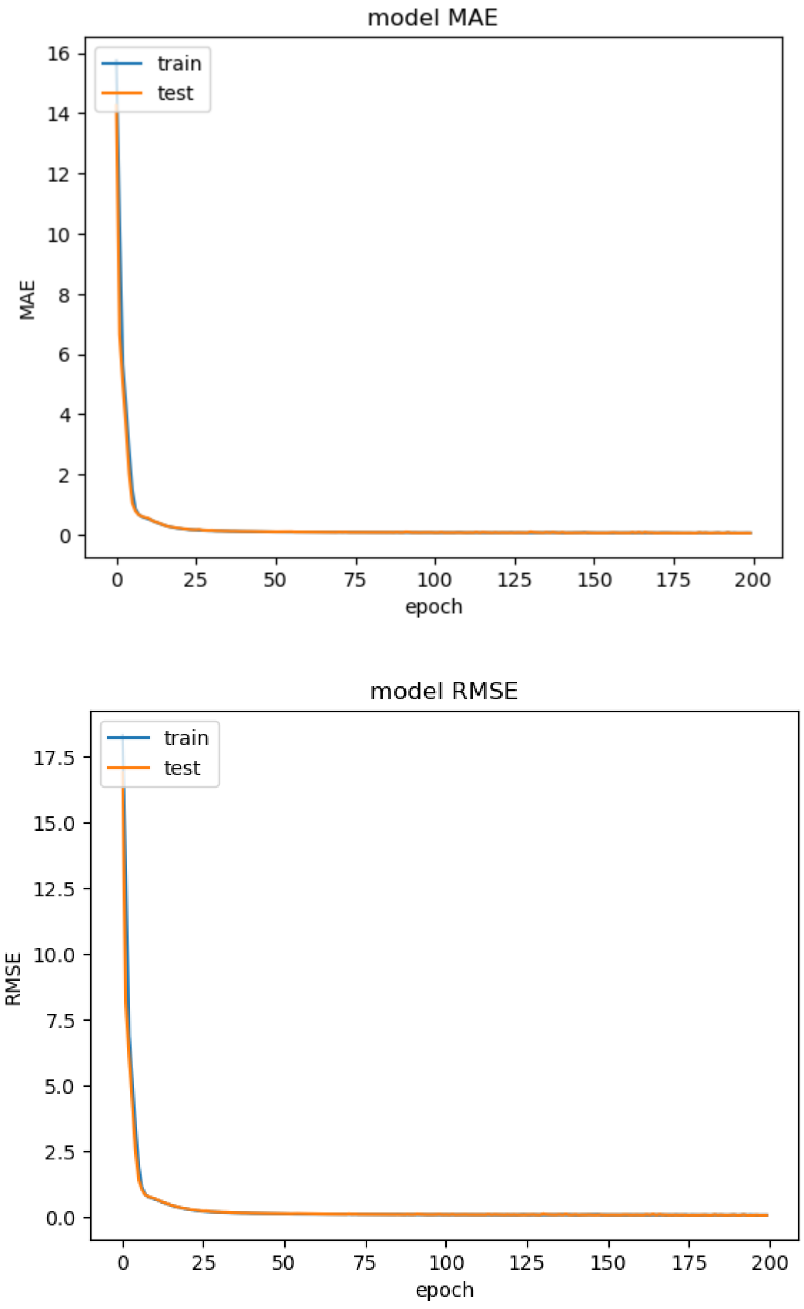
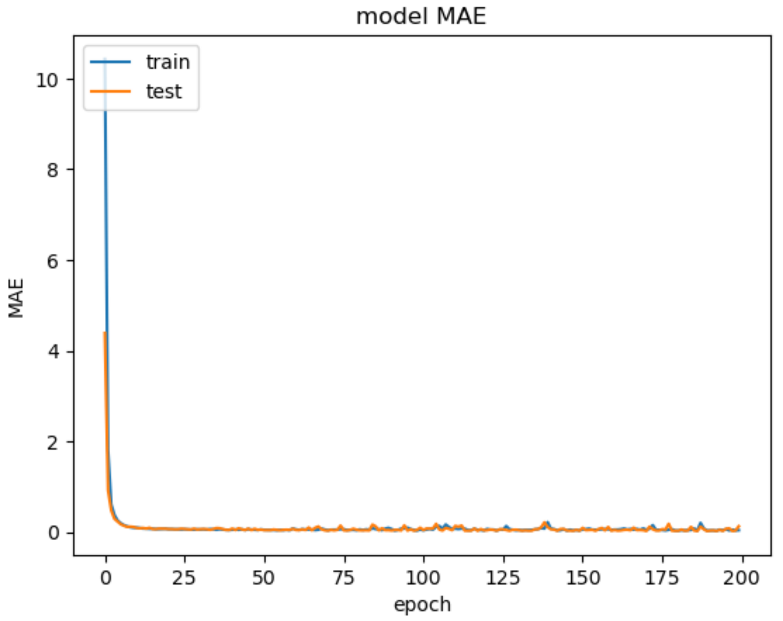


Figure No. 8.3: Model 3 Train and Test metrics

Model 4 Train and test metrics:



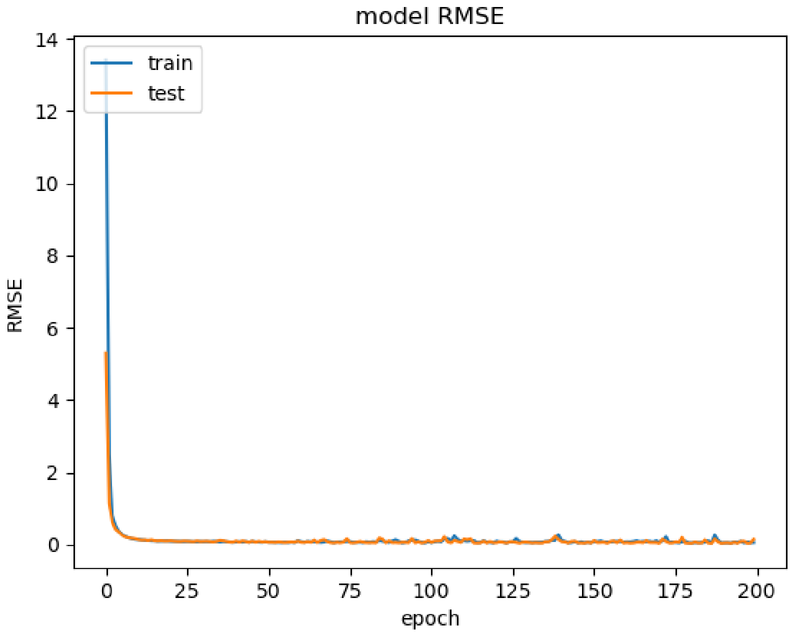


Figure No. 8.4: Model 4 Train and Test metrics

From MAE, RMSE metric data from all 4 models, it is learned that only 25 epochs should be sufficient for learning instead of 200.

Chapter 9: Analysis and Results

Application of different varied DL model configurations performed on dataset and validation metrics of the models are noted to understand the best fit model for a particular design.

Model 1 is taken as baseline model and all other models are compared with baseline model. It can be observed that model 2, MAE/RMSE is 75% less than baseline model. Similarly additional layers and neurons will reduce this to almost zero in model 4, but the cost of building the complex model is very high. Hence, we recommend model 2 for less critical applications like the one that has been tested.

Also, it is observed from chapter 8 results that the Model 4 outperforms rest of the model at epoch 25 or less. Epoch 25 considered the fact that most of them reaches its learning cycle by then.

For Model 1, being simple it took approximately 125 epochs to reach the lowest loss in the model in Figure No. 9.1. At Epoch 25 it performs least compared to all as mentioned in Table 8.1

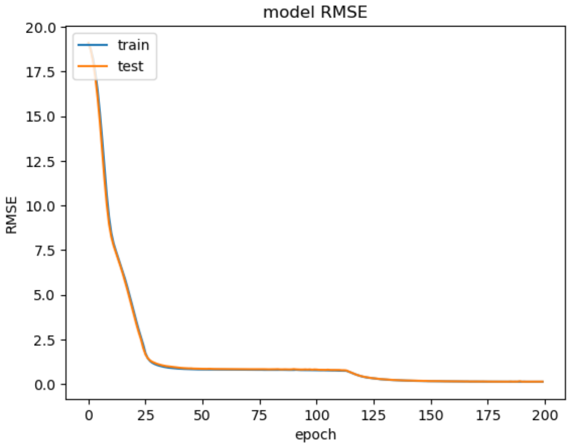


Figure No. 9.1: RMSE metric model 1

For Model 2, being simple and more hidden layer, it took approximately 35 epochs to reach the lowest loss in the model in Figure No. 9.2. At Epoch 25 it performs better than the model 1.

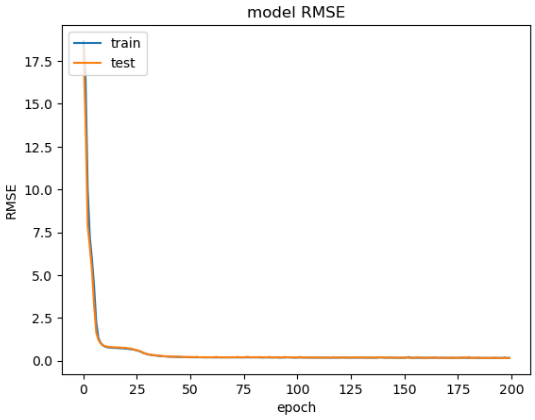


Figure No. 9.2: RMSE metric model 2

For Model 3, being medium and simple it took approximately 25 epochs to reach the lowest loss in the model in Figure No. 9.3

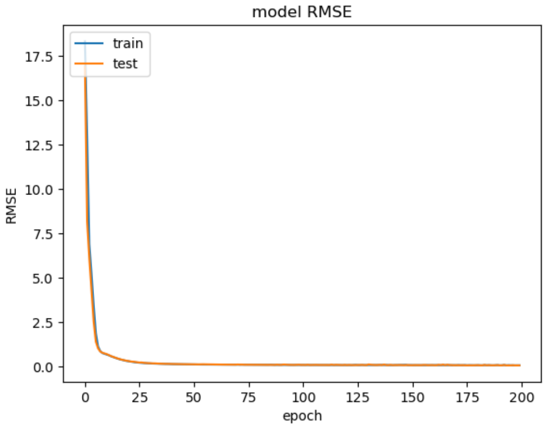


Figure No. 9.3: RMSE metric model 3

For Model 4, being medium and more hidden layer, it took approximately 10 epochs to reach the lowest loss in the model in Figure No. 9.4.

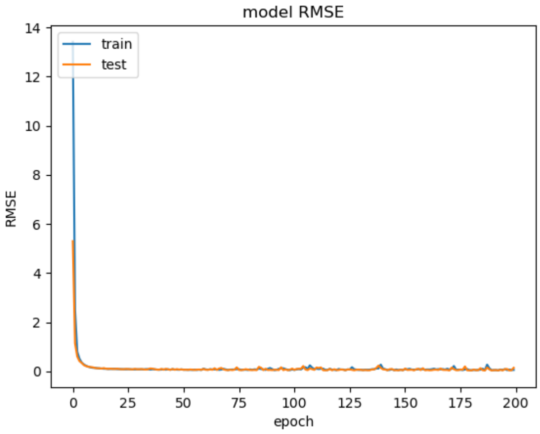


Figure No. 9.4: RMSE metric model 4

The final training and testing converging point or metric value for all models discussed are mentioned in the below Table No. 9.1

Model	Converging Epoch
1	125
2	35
3	25
4	10

Table No. 9.1: Model Vs Converging Epoch

Also, it is evident from Table 9.1 that more complex or deeper the model, it takes less epoch to converge.

Chapter 10: Conclusions and Future Scope

The results from the previous section indicate that deeper and more nodes in the model, train and test converges will happen faster. Model 2 being the best choice for the dataset for depth estimate that we have proposed.

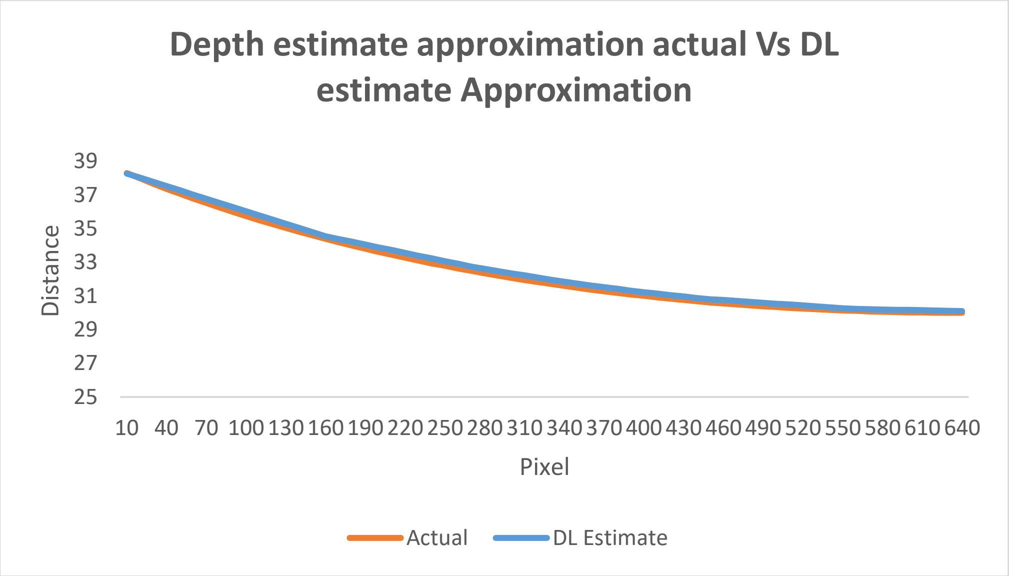


Figure No. 10.1: Actual Vs Estimate with model 2

From the Figure No. 10.1, it is evident that actual vs approximate depth using Deep learning model is significantly small.

As discussed in the chapter 9, then following Table No. 10.1 is used for the selection of the model.

	MAE		**RMSE**	
	Train	Test	Train	Test
Model 1	2.0245	1.8609	2.6184	2.4958
Model 2	0.5060	0.5226	0.6550	0.6694
Model 3	0.1605	0.1675	0.2327	0.2494
Model 4	0.0623	0.0630	0.0837	0.0839

Table No. 10.1: Train and test metrics for varies model configurations at Epoch 25

The end result demonstrates a single-camera Depth Estimation algorithm. You can equip your car with this device to help you avoid collisions. To avoid crashes, a safe distance between the host vehicle and other vehicles must be maintained; our proposed system assesses this distance.

With a higher-resolution camera, the same method can be applied to considerably greater distances. Additionally, the approach can be improved for steep grades and winding routes.

Bibliography

[1] F. Engels, P. Heidenreich, M. Wintermantel, L. Stacker, M. Al Kadi, and A. M. Zoubir, "Automotive Radar Signal Processing: Research Directions and Practical Challenges," *IEEE J. Sel. Top. Signal Process.*, vol. 15, no. 4, pp. 865–878, 2021, doi: 10.1109/JSTSP.2021.3063666.

[2] S. Ul and H. Syed, "Lidar Sensor in Autonomous Vehicles," no. March, 2022.

[3] H. He, S. Zhai, S. Zeng, K. Fu, J. Ruan, and Y. Wang, "Research on Laser Distance Measurement Technology for Remaining Volume of Large Open-pit Material Yard," *IOP Conf. Ser. Earth Environ. Sci.*, vol. 558, no. 2, 2020, doi: 10.1088/1755-1315/558/2/022012.

[4] A. Carullo and M. Parvis, "An ultrasonic sensor for distance measurement in automotive applications," *IEEE Sens. J.*, vol. 1, no. 2, pp. 143–147, 2001, doi: 10.1109/JSEN.2001.936931.

[5] S. Kumar, D. Gupta, and S. Yadav, "Sensor Fusion of Laser & Stereo Vision Camera for Depth Estimation and Obstacle Avoidance," ©*2010 Int. J. Comput. Appl.*, vol. 1, no. 26, pp. 975–8887.

[6] I. Gat, M. Benady, and A. Shashua, "A monocular vision advance warning system for the automotive aftermarket," *SAE Tech. Pap.*, 2005, doi: 10.4271/2005-01-1470.

[7] A. Joglekar, D. Joshi, R. Khemani, S. Nair, and S. Sahare, "Depth Estimation Using Monocular Camera," *Int. J. Comput. Sci. Inf. Technolgies*, vol. 2, no. 4, pp. 1758–1763, 2011.

[8] Y. M. Chiang, N. Z. Hsu, and K. L. Lin, "Driver assistance system based on monocular vision," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 5027 LNAI, no. 1, pp. 1–10, 2008, doi: 10.1007/978-3-540-69052-8_1.

[9] A. Rahman, A. Salam, M. Islam, and P. Sarker, "An image based approach to compute object distance," *Int. J. Comput. Intell. Syst.*, vol. 1, no. 4, pp. 304–312, 2008, doi: 10.1080/18756891.2008.9727627.

[10] N. Yamaguti, S. Oe, and K. Terada, "A Method of Distance Measurement by Using Monocular Camera."

[11] E. Dagan, O. Mano, G. P. Stein, and A. Shashua, “Forward collision warning with a single camera,” *IEEE Intell. Veh. Symp. Proc.*, no. July 2004, pp. 37–42, 2004, doi: 10.1109/ivs.2004.1336352.

[12] P. Alizadeh, “Object Distance Measurement Using a Single Camera for Robotic Applications by Peyman Alizadeh A thesis Submitted in partial fulfillment of the requirements for the degree of Master of Applied Sciences (M A Sc) in Natural Resources Engineering The Facult,” *Object Distance Meas. Using a Single Camera Robot. Appl.*, p. 126, 2015.

[13] S. Diamantas, S. Astaras, and A. Pnevmatikakis, “Depth estimation in still images and videos using a motionless monocular camera,” *IST 2016 - 2016 IEEE Int. Conf. Imaging Syst. Tech. Proc.*, no. March 2018, pp. 129–134, 2016, doi: 10.1109/IST.2016.7738210.

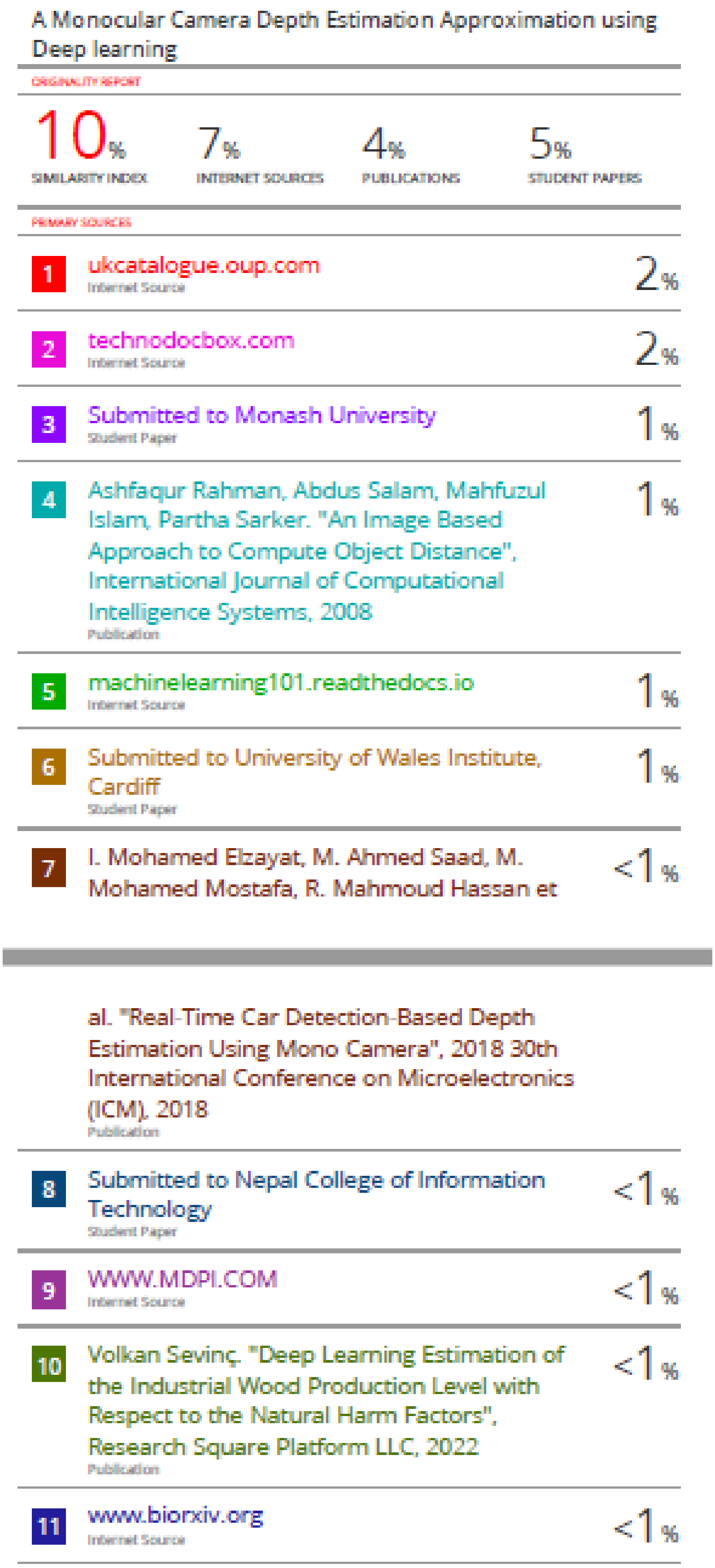
[14] N. K. Chauhan and K. Singh, “A review on conventional machine learning vs deep learning,” *2018 Int. Conf. Comput. Power Commun. Technol. GUCON 2018*, pp. 347–352, 2019, doi: 10.1109/GUCON.2018.8675097.

[15] L. Alzubaidi *et al.*, *Review of deep learning: concepts, CNN architectures, challenges, applications, future directions*, vol. 8, no. 1. Springer International Publishing, 2021.

[16] J. Zhu and Y. Fang, “Learning object-specific distance from a monocular image,” *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2019-Octob, pp. 3838–3847, 2019, doi: 10.1109/ICCV.2019.00394.

Appendix

Plagiarism Report



A Monocular Camera Depth Estimation Approximation using Deep learning

ORIGINALITY REPORT

10%	7%	4%	5%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	ukcatalogue.oup.com Internet Source	2%
2	technodocbox.com Internet Source	2%
3	Submitted to Monash University Student Paper	1%
4	Ashfaqur Rahman, Abdus Salam, Mahfuzul Islam, Partha Sarker. "An Image Based Approach to Compute Object Distance", International Journal of Computational Intelligence Systems, 2008 Publication	1%
5	machinelearning101.readthedocs.io Internet Source	1%
6	Submitted to University of Wales Institute, Cardiff Student Paper	1%
7	I. Mohamed Elzayat, M. Ahmed Saad, M. Mohamed Mostafa, R. Mahmoud Hassan et al. "Real-Time Car Detection-Based Depth Estimation Using Mono Camera", 2018 30th International Conference on Microelectronics (ICM), 2018 Publication	<1%
8	Submitted to Nepal College of Information Technology Student Paper	<1%
9	WWW.MDPI.COM Internet Source	<1%
10	Volkan Sevinç. "Deep Learning Estimation of the Industrial Wood Production Level with Respect to the Natural Harm Factors", Research Square Platform LLC, 2022 Publication	<1%
11	www.biorxiv.org Internet Source	<1%

Publications in conference

Paper submitted and accepted:

Rajanna, *Rashmi Agarwal, Jay Bharateesh Simha, "***A Monocular Camera Depth Estimate Approximation using Deep learning",** IEEE INCOFT 2022 - International Conference on Futuristic Technologies, Submission Date: 25th September 2022, Accepted Date: 1st Oct 2022

A Monocular Camera Depth Estimate Approximation using Deep learning

Rajanna
REVA Academy of Corporate Excellence-RACE
REVA University
Bangalore, India
rajannak.ai01@race.reva.edu.in

Rashmi Agarwal
Reva Academy of Corporate Excellence-RACE
REVA University
Bangalore, India
0000-0003-1778-7519

Jay Bharateesh Simha
Reva Academy of Corporate Excellence-RACE
Reva University
Bangalore, India
jb.simha@reva.edu.in

***Abstract*— In numerous applications, such as collision detection, Robotic handling, Robotic-based manufacturing facility, and Advanced Driver Assistance Systems (ADAS), depth estimation is crucial and the most significant task. Radar, Ultrasonic, Lidar technologies both operate by reflecting radio or sound wave or laser beams respectively. Stereo cameras used for depth estimation are costly and increase the cost of the system. There are a few Monocular camera depth estimate approaches that have evolved using mathematical calculation. One approach uses pixel to depth estimate mapping and the other uses the geometry of the road, the contact point of the vehicle on road, and camera properties. The proposed solution would implement depth detection using a monocular camera with deep learning. The objective of the proposed solution is to detect the depth from a monocular image and calibrate it with the actual depth. The pixel to distance data which derived from camera properties is used to run through varied hidden layers and nodes to conclude the implementation using Deep Learning (DL). Evidently, train and test data learn and converge more quickly on the deeper models which has more hidden layers and nodes.**

Keywords— ADAS, Lidar, Radar, Ultrasonic, Deep Learning, Monocular camera, Depth estimate

I. Introduction

For depth estimate, the most prevalent methods nowadays are based on Radar, Lidar, Ultrasonic, or a combination of these technologies. The radar technology measures metal object reflections and accounts for the Doppler effect to determine relative speed. The Lidar systems estimate distance using laser beams. Radar systems are generally relegated to high-end vehicles due to their prohibitively expensive price tags. Using stereo cameras, the conventional method of depth estimate [1] is limited in range and expensive. Depth estimation using a single camera is challenging because of distortions in the image caused by the camera's perspective. A monocular camera method exists to estimate an item's size, however, it is not very precise and has a tendency to estimate an unknown-sized object incorrectly. A significantly more precise estimate can be derived by considering the road geometry and the vehicle's contact point with the ground. The flatness of the road and the camera's optical axis being perpendicular to the road are prerequisites for this proposed method. In this research, the pixel to distance data which derived from camera properties is used to run through varied hidden layers and nodes to conclude the implementation using deep learning.

II. Literature review

Positive findings have emerged from recent studies of how artificial intelligence can be used in fields like image processing, driverless vehicles, and robotic manufacturing. One study [2][3] discusses how to take depth measurements using a single camera. Train a system to recognize a mapping between an object's pixel height and physical distance by observing and taking advantage of the correlation between these two quantities. The height of each test object in the image is then calculated by using this mapping. Few studies [4]-[6] discuss a further monocular vision system technique for determining range and measuring range. Since only one camera is used to gather data, all distances must be approximated in terms of perspective. Both the size of the truck in the photograph and its location near the bottom of the frame could be indicators. A width-based range estimate has a 30% margin of error because the width of an unidentified vehicle (car, van, truck, etc.) could be anywhere from 1.5 meters to 3 meters. Using the geometry of the road and the vehicle's contact point with the road, a far superior solution was provided. The assumption was a road surface that is flat and a camera with a parallel optical axis to the road surface. In [5], [7], [8] for example, it is suggested that a single forward-facing camera installed on the dashboard be used to perform both in-path and oblique distance calculations.

Deep learning is a subclass of machine learning that comprises neural networks with three or more layers. These neural networks aim to imitate the behavior of the human brain, albeit their ability to "learn" from massive amounts of data falls far short of that of the human brain. The first advantage of deep learning over machine learning is that feature extraction is redundant. Artificial neural networks utilizing deep learning do not require the feature extraction stage. Various studies showed promising results in applying mathematics estimation behind the depth perspective but not many employ deep learning [8] [9]. Therefore, this study addresses this research gap.

III. Problem Statement

The depth estimation technique helps drivers avoid collisions using only one of the cameras often found in modern vehicles. Many different techniques have been devised to help drivers keep a safe distance between their host car and other vehicles on the road. Some approaches make use of the correlation between the actual height of an object and its pixel size. Making use of this connection by training a system to find a mapping between the height of a pixel and the actual distance to an object. The height of each test object in the image is then calculated by using this mapping. Also, a flat road, a perpendicular visual axis to the road, and camera properties are considered. All the methods use a mathematical calculation to estimate the depth. None of them are using contemporary adaptations of deep learning and computer vision models.

IV. Objectives of the Study

Deep learning is the latest trend in the industry and gives breaking results in various fields may it be Text analytics, Natural language processing, computer vision, and so on. DL approach in depth estimation will lead way for fast adaption of the latest models and further improves the potential of industry acceptance. The primary objective of this study is to implement the Depth estimation approximation of a monocular camera system using DL model. This enables the migration of the legacy system to DL model which helps better integrate to AI ecosystem overall. The secondary objective of the study helps in the selection of an optimal DL model to balance the complexity of the model and the cost associated with a dataset of applications that you are intended to use.

V. Project Methodology

Height, focal length, camera tilt, and pixel resolution are all characteristics of cameras. It is important to note that each camera's settings are specific to that camera. The acquired image is in the two-dimensional plane, and the distance can be estimated by determining the relationship between the two-dimensional image and the real three-dimensional view of the image. If an accident is avoided, it will be because the camera was mounted to the automobile and determined the exact distance of the obstacle. Under the assumption of a flat road, a camera is installed so that its optical axis is perpendicular to the road. If we place a point on the road, *Z* meters in front of the camera, that spot will be at a certain vertical position in the image captured by the camera as shown in Fig. 1.

An image plane *(I)* and pinhole *(P)* separated by a focal distance *(f)* make up the schematic pinhole camera shown in Fig. 1. The camera is mounted on the car *(A)* at a certain height *(H)*. The rare end of the car *(B)* is off to the side, at a distance of *Z1*. When viewed from above, the image plane is intersected by the vehicle's point of contact with the road *(Y1)*. Neither the f-number nor the Y-coordinate of the image, which are typically expressed in pixels, are to scale in this illustration.

When a further away car *(C)* makes road contact *(Z2),* its contact point will be projected on its image plane at a lower *(y2)* location than *(y1).*

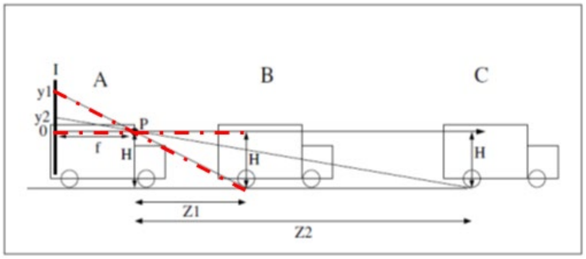


Fig. 1. Schematic diagram of the imaging geometry

The similarity of triangles allows for a straightforward derivation of Equation (1).

$$\frac{Y}{f} = \frac{H}{Z} \tag{1}$$

A. *Flowchart of the proposed methodology*

The workflow in the new approach is shown in Fig. 2. In the new approach, DL is used to compute the depth estimation approximation.

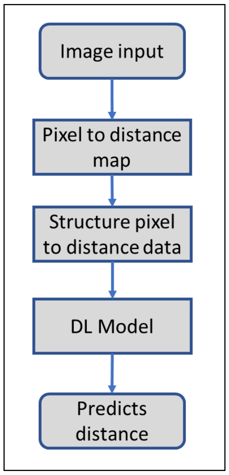


Fig. 2. Proposed methodology

B. *Dataset*

Data for the proposed implementation is simulated using camera parameters like focal length, vertical view angle, horizontal view angle, and pixel density in both vertical and horizontal. The Centre of the typical image as in Fig No. 3 used for data extraction and tabulated in TABLE I. Actual image size is 1280 x 720 pixels and since the 640 x 360-pixel point is the farthest vehicle can move and there is no depth information beyond this point. Also, Oblique distance calculation is represented in Fig. 4 and Equation (2).

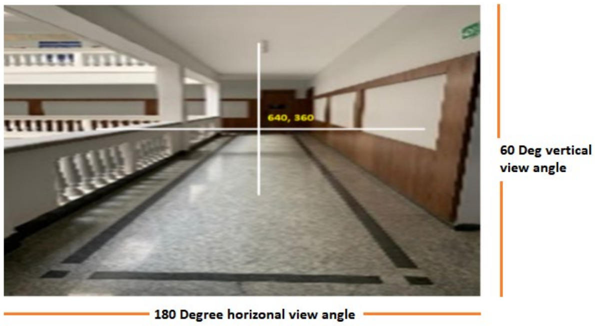


Fig. 3. Pixel to distance calculation

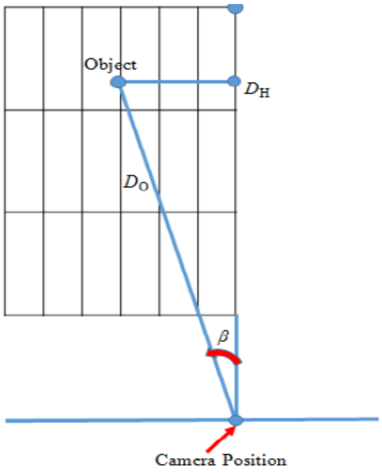


Fig. 4. Oblique distance calculation [7]

$$Do = D_H / Cos(\beta) \qquad (2)$$

Do is the oblique distance, D_H is the straight-line distance from the camera position and *Cos(β)* is the cosine of angle *β* to obtain the oblique distance.

TABLE I. INPUT AND TARGET VARIABLES FOR MODELING

Properties	**Input variables**
Input	Horizontal pixel position
	Vertical pixel position
Output	Depth estimate

C. *Deep learning model configurations*

Many simple to decent complex model has been implemented in order to arrive at the optimal configuration. Four different model configurations as listed in TABLE II are tried, and a comparison is done to arrive at the final solution. Note that the proposed configuration is subjected to train and test datasets split of 70% and 30% respectively. Mean absolute Error (MAE) and Root Mean Squared Error (RMSE) is chosen as loss functions with metric as accuracy. Adam's algorithm is chosen as an optimizer.

TABLE II. MODEL CONFIGURATIONS FOR TEST AND VALIDATION

	No. of nodes at layer 1	**No. of hidden layers**	**No. of nodes in the hidden layer**
Model 1	13	1	10
Model 2	13	4	10
Model 3	50	1	100
Model 4	50	4	100

VI. RESULT AND ANALYSIS

Application of different varied DL model configurations performed on dataset and validation metrics of the models are noted to understand the best fit model for a particular design in Table III. Model 1 is taken as the baseline model and all other models are compared with the baseline model. It can be observed that Model 2, MAE and RMSE are 75% less than the baseline model. Similarly, additional layers and neurons will reduce this to almost zero in Model 4, but the cost of building the complex model is very high. Hence, we recommend Model 2 for less critical applications like the one that has been tested.

TABLE III. TRAIN AND TEST METRICS FOR VARIES MODEL CONFIGURATIONS AT EPOCH 25

	MAE		**RMSE**	
	Train	***Test***	***Train***	***Test***
Model 1	2.0245	1.8609	2.6184	2.4958
Model 2	0.5060	0.5226	0.6550	0.6694
Model 3	0.1605	0.1675	0.2327	0.2494
Model 4	0.0623	0.0630	0.0837	0.0839

VII. CONCLUSION

The results from the previous section indicate that deeper and more nodes in the model, train and test converges will happen faster. Model 2 being the best choice for the dataset for depth estimate that we have proposed.

The end result demonstrates a single-camera Depth Estimation algorithm. One can equip car with this device to help you avoid collisions. To avoid crashes, a safe distance between the host vehicle and other vehicles must be maintained; The proposed system assesses this distance. With a higher-resolution camera, the same method can be applied to considerably greater distances. Additionally, the approach can be improved for steep grades and winding routes.

REFERENCES

[1] S. Kumar, D. Gupta, and S. Yadav, "Sensor Fusion of Laser & Stereo Vision Camera for Depth Estimation and Obstacle Avoidance," ©2010 Int. J. Comput. Appl., vol. 1, no. 26, pp. 975–8887.

[2] A. Rahman, A. Salam, M. Islam, and P. Sarker, "An image based approach to compute object distance," Int. J. Comput. Intell. Syst., vol. 1, no. 4, pp. 304–312, 2008, doi: 10.1080/18756891.2008.9727627.

[3] N. Yamaguti, S. Oe, and K. Terada, "A Method of Distance Measurement by Using Monocular Camera.

[4] I. Gat, M. Benady, and A. Shashua, "A monocular vision advance warning system for the automotive aftermarket," SAE Tech. Pap., 2005, doi: 10.4271/2005-01-1470.

[5] A. Joglekar, D. Joshi, R. Khemani, S. Nair, and S. Sahare, "Depth Estimation Using Monocular Camera," Int. J. Comput. Sci. Inf. Technolgies, vol. 2, no. 4, pp. 1758–1763, 2011.

[6] E. Dagan, O. Mano, G. P. Stein, and A. Shashua, "Forward collision warning with a single camera," IEEE Intell. Veh. Symp. Proc., no. July 2004, pp. 37–42, 2004, doi: 10.1109/ivs.2004.1336352.

[7] P. Alizadeh, "Object Distance Measurement Using a Single Camera for Robotic Applications by Peyman Alizadeh A thesis Submitted in partial fulfillment of the requirements for the degree of Master of Applied Sciences (M A Sc) in Natural Resources Engineering The Facult," Object Distance Meas. Using a Single Camera Robot. Appl., p. 126, 2015.

[8] S. Diamantas, S. Astaras, and A. Pnevmatikakis, "Depth estimation in still images and videos using a motionless monocular camera," IST 2016 - 2016 IEEE Int. Conf. Imaging Syst. Tech. Proc., no. March 2018, pp. 129–134, 2016, doi: 10.1109/IST.2016.7738210.

[9] N. K. Chauhan and K. Singh, "A review on conventional machine learning vs deep learning," 2018 Int. Conf. Comput. Power Commun. Technol. GUCON 2018, pp. 347–352, 2019, doi: 10.1109/GUCON.2018.8675097.