ATCSJ
Applied Technology and Computing Science Journal
Open Access
p-ISSN 2621-4458 | e-ISSN 2621-4474

UNUSA

Available online to journal2.unusa.ac.id

UNUSA
S4-Accredited – SK No.200/M/KPT/2021
Journal Page is available to https://journal2.unusa.ac.id/index.php/ATCSJ

Vehicle Safe Distance Detection System Based On Image Processing As Accident Prevention With Faster R-CNN Method

Agus Khumaidi[1*], Elok A. Candra[2], Perwi Darmajanti[3], Ivan A. Septiadi[4], Sryang T. Sarena[5]

[1,2] Teknik Otomasi, Teknik Kelistrikan Kapal, Politeknik Perkapalan Negeri Surabaya, Indonesia
Jl. Teknik Kimia, Kampus ITS, Keputih Sukolilo, Surabaya
[1*]aguskhumaidi@ppns.ac.id

Article history:

Received 13 January 2023
Revised 22 January 2023
Accepted 22 January 2023
Available online 22 January 2023

Keywords:

Faster R-CNN,
Stereo Vision,
Mono Vision,
Linear Regression,
Object Detection.

Abstract

Numerous victims and huge economic and social losses have resulted from the escalating number of traffic accidents. From these issues, a technique to create a camera capable of detecting vehicles going around the driver using the Faster R-CNN method and calculating the vehicle's distance using the Stereo Vision and Mono Vision methods was discovered. The determination of safe distance between these cars is determined by the speed of the driver's vehicle, with the LED and buzzer warning system activating when the parameters are met. Based on the results of object detection experiments utilizing the Faster R-CNN, the model's success rate in identifying and classifying objects had an average success rate of 83.33 percent across 35 object situations examined from different perspectives. The success rates for distance estimates utilizing the Stereo Vision and Mono Vision methods with the Linear Regression equation were 98.84% and 98.10%, respectively.

I. INTRODUCTION

In Indonesia, traffic accidents are one of the leading causes of mortality. With a relatively high number of victims, economic losses (material losses), and social repercussions that are not negligible, numerous preventative initiatives have been made to improve traffic by incorporating several [1]. The Central Statistics Agency reports that the number of accidents in Indonesia in 2019 reached 116,411 (Central Statistics Agency, 2022). In numerous locations, the availability and building of traffic amenities, specifically toll highways, have been observed. Where traffic is facilitated by a motorway in emerging regions. Data reveals that 343 incidents happened on the Trans Java toll road between December 2018 and January 2019, with 246 accidents caused by driver carelessness, 89 accidents caused by cars, and 8 accidents caused by the environment [2] [3].

[1*] Corresponding author

ISSN 2621-4474 (online) 2621-4458 (print) © 2022 The Authors. Published by UNUSA

DOI: https://doi.org/10.33086/atcsj.v5i2.3804.

According to these statistics, the most major cause of accidents is the driver's lack of awareness or focus while driving. "Image Processing Based Vehicle Safe Distance Detection System as Accident Prevention with the Faster R-CNN Method" is an innovation developed by the author to limit the frequency of accidents on toll roads caused by driver irresponsibility. This study uses a camera that captures images similarly to the human eye. It is processed with a specific classification using OpenCV so that the video may be processed in real-time, categorized into many objects using the Faster R-CNN technique, and distances between objects and the camera can be predicted using the Stereo Vision and Mono Vision methods.

II. FASTER REGION CONVOLUTIONAL NEURAL NETWORK (FASTER R-CNN)

Faster Region-based Convolutional Neural Networks (Faster R-CNN) is a detection technique whose primary architecture is Fast R-CNN and RPN. This method is a modification of the Fast R-CNN by replacing its selective search part with RPN. RPN is a neural network that substitutes the role search to submit region. RPN generates certain bounding boxes, with each box having two probability scores, indicating whether an item exists at that position. These areas will serve as inputs for comparable designs, such as Fast R-CNN. Using RPN to replace selective search can drastically lower the computing resources required to make the entire model viable and trainable from beginning to finish [4] [5]. Figure 1 depicts the architecture of the Faster R-CNN algorithm.

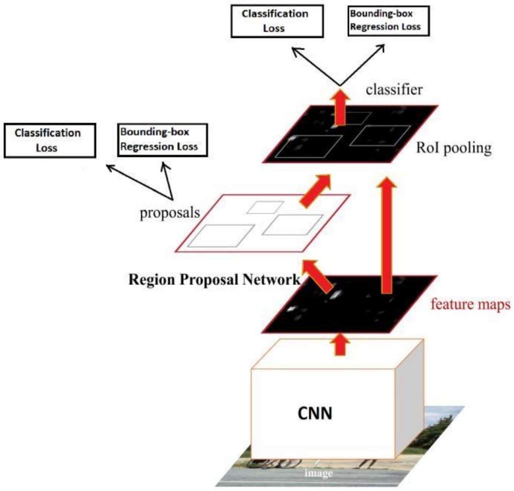

Figure 1. Architecture *Faster R-CNN* [5].

Faster R-CNN is divided into 2 (two) important parts, namely:

1. Region Proposal Network (RPN)
   Region Proposal Network (RPN) RPN is a process that aims to explore possibilities for the location of objects in the image that is inserted quickly. The object location entered has object constraints from that region identified, namely the Region of Interest (ROI). Inputs used on The ROI layer is a feature map which is the output of CNN with multiple convolution layers and max pooling layers. In RPN, initially, the input image is processed in a neural convolution to produce a feature map consisting of 6 (six) sections, viz determination of objects and non-objects with a value of 0-1, the coordinates of the value x and y, as well as the weight and height values of the bounding box. Sliding windows are placed on each feature map with size $N \times N$, accordingly with each anchor sliding window formed. Every anchor has the same center point but has aspect ratios and different calling factors.
2. Classifier
   The classifier is utilized to categorize the ROI detected by the RPN into classes using CNN.

A. ***Stereo Vision***

A stereo vision system (Stereo Vision) is a field concerned with detecting the three-dimensional structure of a scene using two or more digital pictures captured from varying perspectives. A stereo camera consists of two identical cameras positioned in the horizontal and vertical planes in a straight line. When the item is at a point of view overlap between the two cameras, distance measurements are made [6] [7] [8].

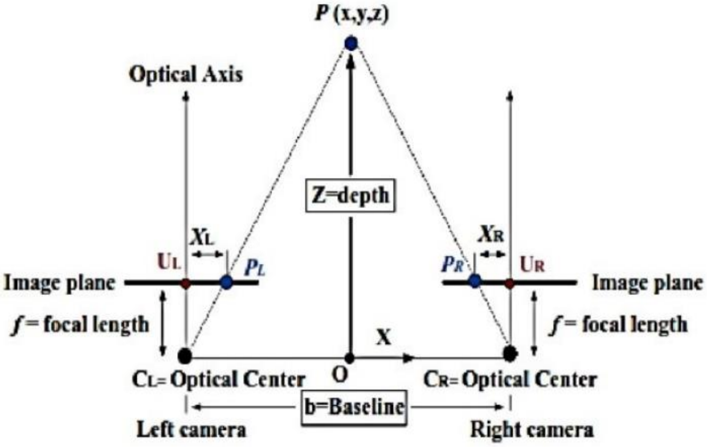

Figure 2. Epipolar Geometry with Aligned Optical Axes [6].

Figure 5 is a camera stereo coordinate system that is assumed to be in the midpoint between the coordinate systems of the right camera and the left camera. From the concept of congruence triangles ΔPC_LC_R and ΔP_LP_R in Figure 5, the stereo coordinate system of the two cameras can be written in Equation 1 and 2.

$$\frac{b}{Z} = \frac{(b + X_R) - X_L}{Z - f} \tag{1}$$

$$Z = \frac{b \times f}{X_L - X_R} = \frac{b \times f}{d} \tag{2}$$

With,

$d = (X_L - X_R)$ = Disparity

X_L = Coordinate$-x$ on the left image

X_R = Coordinate -y on the right image

b = Baseline length (distance between the optical axes of the two cameras)

f = Camera Focus Length

B. ***Mono Vision***

Regression is a statistical technique used to model the value of one variable (independent variable) based on the value of another variable (dependent variable). Linear Regression is the most straightforward kind of Regression. Straight lines are used to represent linear regression. Equation 3 represents the equation for linear regression.

$$Y = a + bX \tag{3}$$

Information:

Y = Dependent variable

X = Independent variable

a = Constan

b = Regression coefficient

Where to find the value of a and b, namely by using Equations 4 and 5.

$$b = \frac{n(\Sigma XY) - (\Sigma X)(\Sigma Y)}{n(\Sigma X^2) - (\Sigma X)^2} \tag{4}$$

$$a = \frac{\Sigma Y - \mathrm{b}(\Sigma X)}{n} \tag{5}$$

III. METHOD

This section explains the experimental design, instruments, methods of data collecting, and control types. This research addresses the issue of how the system can detect passing automobiles using the Faster R-CNN approach and estimate their distance using the Mono Vision method. Then, after the system recognizes the car, the vehicle's distance can be integrated into the warning system via displays, buzzers, and LEDs. The purpose of this research is to develop tools capable of estimating the speed of vehicles passing a car driver in order to improve driver vigilance. Figure 3 is a flowchart that systematically depicts the research's steps.

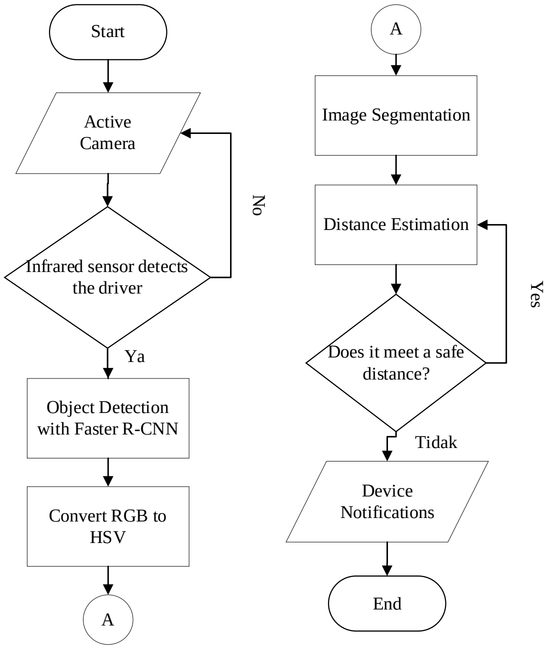

Figure 3. System Flowchart

The parameters for estimating safe distances are dependent on the speed of the user's vehicle, as indicated in Table 1, which is the authors' expansion of the speed data and safe distance [9] [10].

Table 1. Vehicle Speed and Safe Distance

Speed	Minimum Distance	Alert Distance	Safe Distance
≤64 km/hr	≤40 mater	50 mater	60 mater
65-69 km/ hr	45 mater	55 mater	65 mater
70-74 km/ hr	50 mater	60 mater	70 mater
75-79 km/ hr	55 mater	65 mater	75 mater
80-84 km/ hr	60 mater	70 mater	80 mater
85-89 km/ hr	65 mater	75 mater	85 mater
90-94 km/ hr	70 mater	80 mater	90 mater
95-99 km/ hr	75 mater	85 mater	95 mater
≥100 km/ hr	80 mater	90 mater	≥100 mater

If a vehicle speed between 60 and 100 km/h is detected and the minimal space between the driver's car and other surrounding vehicles has been reached, the warning system will activate. The warning system consists of a buzzer and LED flash sound alert. At a minimum distance and alert, the LED's indicator is a red light, which shuts off when the distance is safe.

This study addresses initiatives to lower accident rates, particularly on toll roads. The toll road is a motorway with a minimum speed of 60 kilometers per hour and a maximum speed of 100 kilometers per hour. Due to the excessive speed and proximity of vehicles, this circumstance frequently results in many collisions. How the system can detect things using Faster R-CNN [11] and estimate the distance between objects and cameras using Stereo Vision and Mono Visio [12] with the Linear Regression method [13] is the statement of the research challenge. The system employs the Faster R-CNN algorithm. Figure.4 depicts the procedures required for object detection.

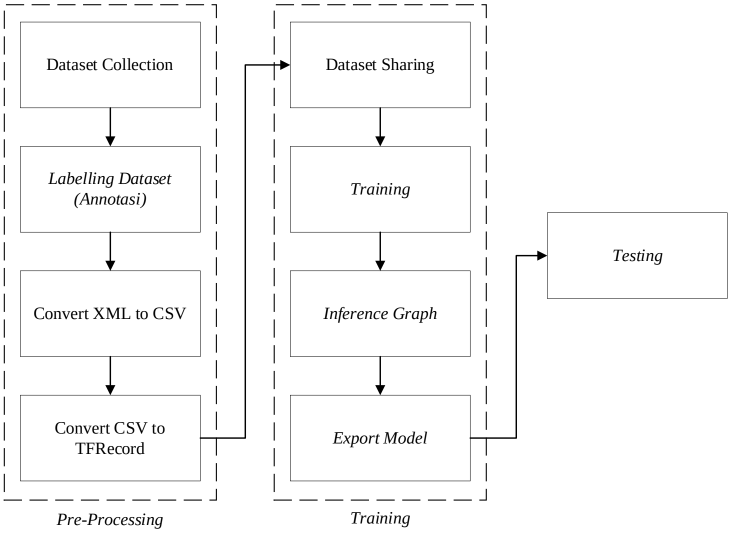

Figure 4. Stages of Object Detection with Faster R-CNN

The first step in the Pre-Processing phase is data collection, which is accomplished by capturing an image of the item with a camera or image collection. The image is subdivided into the following vehicle classes:

(1) Sedan, (2) SUV, (3) Minibus, (4) Truck, (5) Bus, and (6) Pickup. In addition, the collected dataset for each class Annotation or labelling is assigned to each image using Labeling. This image labelling is used to offer information regarding the location of the required '.xml'-formatted picture. The label gives the image a box border and assigns a name image to each class. The labelled data must then be translated from '.xml' to '.csv' and separated into train and test files. After executing the mapping, the.csv data is transformed such that TensorFlow can read it using TFRecord. This conversion is performed to convert the dataset to binary format for optimal training. Following the completion of the dataset preprocessing phase, the data enters the Faster R-CNN training process, which seeks to train 38 image data. When the training process is completed, the results will display the calculation of the loss value and the required training time to acquire the loss value in order to generate a graph of the training dataset's results and extract the data model to classify picture data. Stage Lastly, data is tested by entering a test image. The machine will then read the data model generated by training and proceed with object detection.

IV. RESULTS AND DISCUSSIONS

A. Formation of CNN Architecture

CNN (Convolutional Neural Network) is the primary component of Faster R-CNN. CNN will process the first visual input it receives first. CNN's general procedure consists of three stages: pre-processing, processing, and classification. The pre-processing procedure includes two steps: the generation of datasets and their conversion to grayscale. The second step of the processing procedure involves image convolution, picture dimension reduction, max pooling, and softmax. Classification, the third procedure, determines the output. Figure 5 depicts the design of the Faster R-CNN.

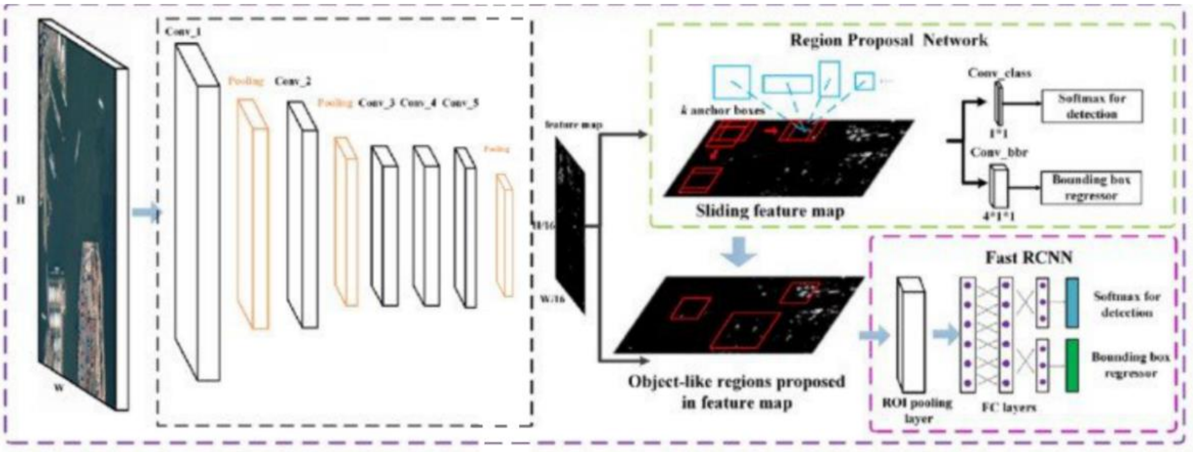

Figure 5. Arsitektur *Faster R-CNN*

The architectural image of CNN on Faster R-CNN can be seen in Figure 6.

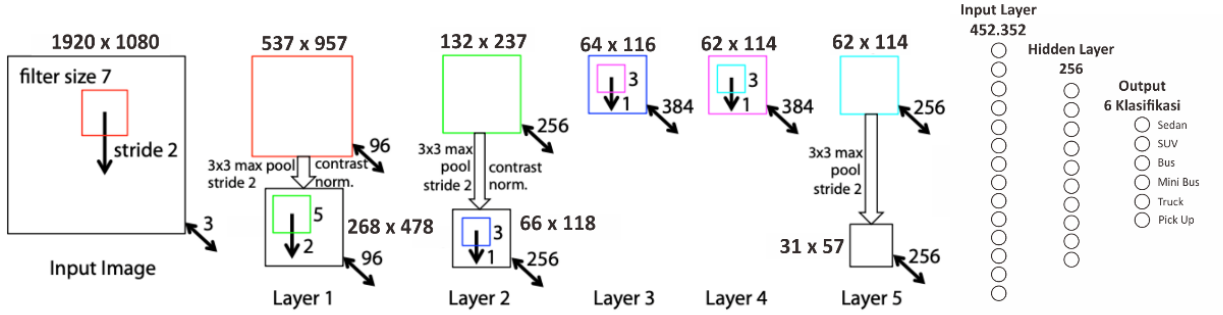

Figure 6. Arsitektur *Faster R-CNN*

This research employs an input image with a resolution of 1080 by 1920 pixels. The RGB image will be converted to grayscale at the specified resolution. The input image will then be subjected to convolution four times and max pooling three times. The result of convolution and max pooling is a feature map that will be transmitted to Region Proposal Network (RPN). In RPN, there is an object classification for generating a proposal object that provides two probabilities as to whether or not an item exists in the specified anchor, as well as a bounding box regressor for changing the bounding box limits to fit the objects inside.

In addition, the object of the proposal generated by the RPN is projected to feature maps generated by CNN and merged ROI (Region of Interest) in order to extract the feature vector corresponding to each object proposal. This procedure will generate layers with input values of 452,352 and hidden layers with 256 neurons. At the final stage, feature maps that have been inputted to fully connected layers will be split into two branches: multiclass classification, which uses multiple layers of convolution and softmax to classify the appropriate proposal object into one of the class categories, and a perfecting bounding box regressor that matches the bounding box boundaries to the objects inside. So that this procedure will generate six object classes, including sedan, SUV, bus, minibus, truck, and pick-up.

B. Conversion RGB to Grayscale

The dataset initially consists of RGB-formatted data. CNN feed is one of the primary components of Faster R-design. The captured dataset will be turned into a grayscale image processing procedure at this point, as CNN can only operate on grayscale images. Figure 7 is a converted RGB to grayscale image.

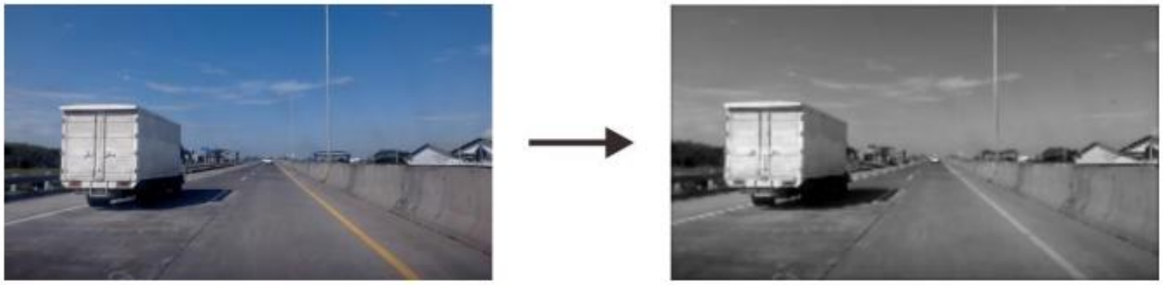

Figure 7. Conversion of RGB to Grayscale

The conversion of RGB photos to grayscale is performed automatically. This grayscale image contains a level of grey. Following the conversion procedure, the input image will be subjected to image convolution.

C. Kernel Convolution Stage and Max Pooling

Convolution is one of the image filtering techniques; in this work, image convolution was performed five times using kernel convolutions measuring 7x7, 5x5, and 3x3 on 1080x1920 greyscale images. The max-pooling of images is a component of the image reduction phase. Image simplification with max pooling by taking three times the most significant value in the axb matrix with a 3x3 filter.

Table 2. Image Convolution Stage and Max Pooling

Original Image	Convolution 1 7×7×96 *stride* 2	*Max Pooling* 1 3×3×96 *stride* 2	Convolution 2 5×5×256 *stride* 2
 1080 × 1920	 537 × 957	 268 × 478	 132 × 237

Max pooling 2 3×3×256 *stride* 2	Convolution 3 3×3×384	Convolution 4 3×3×384	*Max pooling 3* 3×3×256
66 × 118	64 × 116	62 × 114	31 × 57

D. *Dataset*

The collection of datasets is obtained by collecting information on the many sorts of cars that use the toll road. There are 1,000 photos with the '.jpg' file extension included in the datasets. In addition, annotation or labelling of photographs is performed. The annotation procedure generates a file with the '.xml' extension for each dataset file. This file contains a representation of each file in the dataset. After annotation, the dataset is separated into 80% train data and 20% test data.

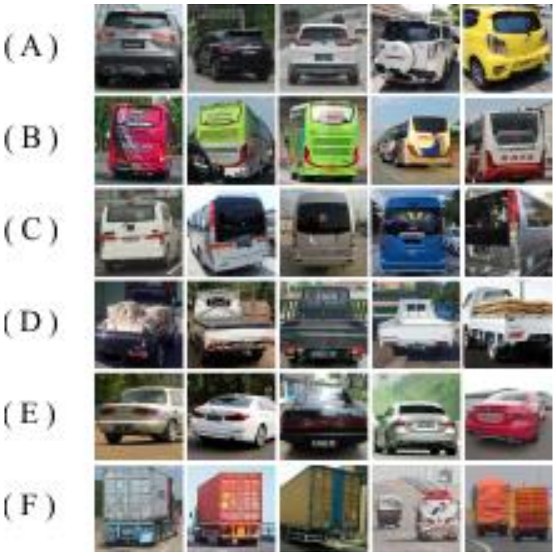

Figure 8. *Dataset Class* (A) SUV (B) Bus (C) Mini Bus (D) Pick Up (E) Sedan (F) Truck

E. *Training*

The.csv train and test data files are converted to TFRecord (TensorFlow Record) files. The TFRecord file is utilised during the training procedure. The training procedure employs the Faster R-CNN model on the Google Collaboratory platform. Google Collaboratory is utilized to aid the training process, which takes a substantial amount of computation. Before training, GPUs connected to Google Drive were used to configure runtime settings. Table 3 displays the loss results from Faster R-CNN dataset training.

Table 3. Dataset Training Results with Faster R-CNN

Step	***Loss***	***Box Classifier***		***RPN***	
		Classification Loss	***Localization Loss***	***Localization Loss***	***Objectness Loss***
20.000	0.01118	0.00213	0.1178	0.02567	0.00552

Step	***Loss***	***Box Classifier***		***RPN***	
		Classification Loss	***Localization Loss***	***Localization Loss***	***Objectness Loss***
40.000	0.02809	0.00061	0.00996	0.0172	0.00031
60.000	0.03551	0.01182	0.01832	0.00338	0.00023
80.000	0.02073	0.00372	0.01181	0.00192	0.00327
100.000	0.02239	0.01596	0.00561	0.00018	0.00063
120.000	0.05333	0.03808	0.00958	0.00551	0.00001
140.000	0.06831	0.00436	0.00151	0.00372	0.00131
160.000	0.06453	0.01696	0.00874	0.00240	0.0002
180.000	0.06022	0.04479	0.01141	0.00237	0.00164
200.000	0.02541	0.01268	0.00920	0.00349	0.00002

During the training process, the system records all processes that occur and saves the file in '.ckpt' format and will stop at step 200,000. The last step will be converted into a training model in the protobuf format with the '.pb' file extension. Files with the protobuf format are used as a model for detection.

F. Inference Graph Tensorboard

Tensorboard is used because the Neural Network is a process known as a black box, where it cannot be observed in detail what processes occur in the system. The training process is carried out in 200,000 steps and produces a Total Loss in the final step of 0.02. The Total Loss graph is shown in Figure 7.

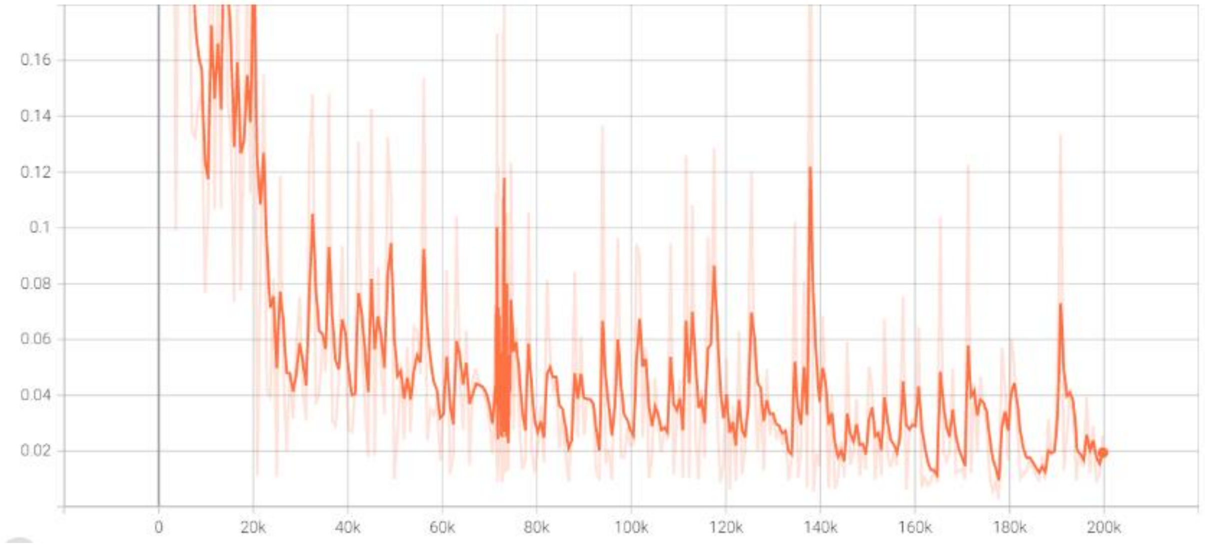

Figure 9. Total Loss chart

G. Object Detection Testing

There are six classes of detection: sedan, SUV, minibus, bus, pickup, and truck. The accuracy of this object detection is evaluated using a variety of driving conditions and object distances and angles. Where researchers conducted tests with the windshield-mounted camera and examined the accuracy of this detection on 28 data samples comprising up to 35 objects.

Table 4. Object Detection Testing

Detection Results	**Number of Objects**	**Percentage**

Fail	5	16.67%
Succeed	30	83.33%

Based on the test results in Table 4, of the 35 objects that have been detected, the average success rate of the system in detecting it is 83.33% with an error rate of 16.67%.

H. Distance Estimation Test

1. Stereo Vision

The distance test estimates the distance between the driver's car and the vehicle in front of it using the Stereo Vision method. The test makes use of the difference between the object's x-coordinate in the left camera image and its y-coordinate in the right camera image. Based on the test findings presented in Table 5, the system's Stereo Vision Method distance estimation accuracy is 98.84%.

Tabel 5. Distance Estimation Accuracy Test Results with Stereo Vision

No	**Detection Distance**	**Actual distance**	**Difference**	***%Error***
1	2.68	2.75	0.07	2.5%
2	2.84	2.85	0.01	0.3
3	3.93	3.85	0.08	2%
4	4.02	4.05	0.03	0.7%
5	4.84	4.90	0.06	1.2%
6	6.46	6.50	0.06	0.6%
7	7.44	7.50	0.06	0.8%
AVERAGE ERROR VALUE				1.16%

2. Mono Vision

Using Linear Regression, testing the distance with the Stereo Vision method to estimate the distance between the driver's car and the vehicle ahead. Based on the test findings presented in Table 6, it is known that the average error value for estimating distances is 1.23 percent, or alternatively, that the system's accuracy rate when predicting object distances using the Mono Vision Method is 98.11 percent.

Table 6. Distance Estimation Accuracy Test Results with Mono Vision

No.	**Detection Distance**	**Actual distance**	**Difference**	***%Error***
1	1.88	2.00	0.12	6%
2	3.26	3.30	0.04	1.2%
3	4.67	4.75	0.08	1.6%
4	5.88	6.00	0.12	2%
5	6.93	6.95	0.02	0.3%
6	7.27	7.25	0.02	0.3%
7	8.26	8.35	0.09	1.07%
AVERAGE ERROR VALUE				1.88%

V. Conclusions And Recommendations

This study divides the system into three major stages: Object Detection, Distance Estimation, and Warning System. The success rate of the system in detecting objects using the Faster R-CNN approach is 83.33 percent and the error rate is 16.67 percent based on 35 conditions of objects tested from various angles and distances. The distance estimate techniques of Stereo Vision and Mono Vision have been integrated into the system. The Stereo Vision approach estimates distances using the coordinate systems of the right and left cameras, whereas the Mono Vision method employs the Linear Regression equation. The success percentage of the system for determining the object's distance from the vehicle is 98.84% for Stereo Vision and 98.11 % for Mono Vision. The warning system in the form of an LED and a buzzer activates when a minimal space between the driver's car and the vehicle in front of it is recognized.

VI. REFERENCES

[1] A. H. W. Fahza, “Analisis Daerah Rawan Kecelakaan Lalu Lintas pada Ruas Jalan Tol Surabaya-Gempol,” *Jurnal Teknik ITS*, vol. 8, no. 1, 2019.

[2] B. P. Statistika, “Jumlah Kecelakaan, Korban Mati, Luka Berat,,” Badan Pusat Statistika, 18 Agustus 2022. [Online]. Available: https://www.bps.go.id/indicator/17/513/1/jumlah-kecelakaan-korban-matiluka-berat-luka-ringan-dan-kerugian-materi.html. [Diakses 2022 Agustus 2022].

[3] B. P. J. Tol, “Peran BPJT dalam Mengantisipasi Perkembangan Jalan Tol Trans Jawa dari Aspek Ekonomi, Sosial, dan Standar Pelayana,” Badan Pengatur Jalan Tol, 2020. [Online]. Available: https://balitbanghub.dephub.go.id/file/222. [Diakses 15 Agustus 2022].

[4] S. Megawan, “Face Spoofing Detection Using Faster R-CNN with Resnet50 Architecture on Video,” *Jurnal Nasional Teknik Elektro dan Teknologi Informasi*, vol. 9, no. 3, pp. 261 - 267, 2020.

[5] A. Khumaidi, “Welding Defect Classification Based on Convolution Neural Network (CNN) and Gaussian Kernel,” dalam *Intelligent Technology and Its Applications (ISITIA) 2017*, Surabaya, Indonesia, 2017.

[6] S. Kaul, Region Based Convolutional Neural Networks for Object Detection and Recognition in Adas Application, United States: The University of Texas at Arlington, 2017.

[7] T. Urip, Pengukuran Jarak Objek Pejalan Kaki Terhadap Kamera Menggunakan Kamera Stereo Terkalibrasi dengan Segmentasi Histogram of Oriented Gradient, Semarang: Universitas Diponegoro, 2017.

[8] I. Marzuqi, “Segmentasi dan Estimasi Jarak Bola dengan Robot Menggunakan Stereo Vision,” dalam *5th Indonesian Symposium on Robotic Systems and Control*, Bandung, Universitas Pendidikan Indonesia, 2017.

[9] F. P. A. Rais Bastomi, “Object Detection and Distance Estimation Tool for Blind People Using Convolutional Methods with Stereovision,” dalam *2019 International Symposium on Electronics and Smart Devices (ISESD)*, Bandung, Indonesia, 2019.

[10] P. R. Indonesia, Peraturan Pemerintah Republik Indonesia No. 43 Tahun 1993 Tentang Prasarana dan Lalu Lintas Jalan, Presiden Republik Indonesia, 1993.

[11] Otoklik, “Segini Jarak Aman Berkendara Menurut Aturan yang Berlaku,” otoklix.com, 19 November 2022. [Online]. Available: https://otoklix.com/blog/jarak-aman-berkendara. [Diakses 29 November 2022].

[12] Y. N. Deta, Klasifikasi Jenis Lubang Kerusakan Jalan Menggunakan Faster Region Convolutional Neural Network dengan Pengambilan Data Secara Vertikal dan Horizontal, Sumatera Utara, Indonesia: Universitas Sumatera Utara, 2021.

[13] F. M. Dirgantara, “Object Distance Measurement System Using Monocular Camera on Vehicle,” dalam *2019 6th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI)*, Bandung, Indonesia, 2019.

[14] P. P. Adikara, “Regresi linier berbasis clustering untuk deteksi dan estimasi halangan pada smart wheelchair,” *Jurnal Ilmiah Teknologi Sistem Informasi*, vol. 3, no. 1, pp. 11-16, 2016.