UDC 004.93'1:004.89
DOI: 10.15587/1729-4061.2025.339774

DEVELOPMENT OF A DISTANCE MEASUREMENT MODEL USING A MAGNIFICATION APPROACH AND MODIFICATION OF THE YOLOV3 ARCHITECTURE

Herdianto Herdianto
Doctoral of Computer Science*
Poltak Sihombing
Correspondence author
Doctor of Computer Science, Professor*
E-mail: poltak@usu.ac.id
Fahmi Fahmi
Doctoral Doctor of Electrical Engineering, Professor
Department of Electrical Engineering**
Tulus Tulus
Doctor of Philosophy, Professor
Department of Mathematics**
*Department of Computer Science**
**Universitas Sumatera Utara
Dr. T. Mansur str., 9, Padang bulan,
North Sumatera, Indonesia, 20155

Determining the best model for measuring object distance, the appropriate formula for that model, and modifying the YOLOv3 architecture is the focus of this research. This was done to address the problem of object distance measurement errors using a monocular camera. In this study, the researchers used a magnification approach and modified the YOLOv3 architecture, which was then named Hybrid Dist – YOLOv3. The proposed distance measurement model does not use camera height and camera shift distance variables, so it can still measure objects that are higher than the camera height and the measurement time is faster. The only variable in the measured distance formula is the change in object image height. As for modifications to the YOLOv3 architecture, there are two types of training and test data: initial measurement data and from KITTI. The training data from the initial measurements consisted of three classes, namely person, bottle, and jerrycan, with 24, 10, and 10 samples, respectively. The detection accuracy at mAP0.50 is 0.994, 1.1, with absolute measurement error values (εA) of –0.274, –0.153, and –0.163. For the training data from KITTI, there are three object classes, namely pedestrian, car, and truck, with 1150, 7682, and 318 samples, respectively. From the tests conducted, the εA values for the pedestrian, car, and truck classes show an improvement from the previous study, which were originally 1.75, 2.49, and 4.63, to 1.37, 2.25, and 3.74. The results of this research can be applied in the automotive industry to driver assistance systems (DAS), soccer robots, or similar systems that require distance measurement

Keywords: detection object, distance estimation, DAS, monocular camera, YOLO, Hybrid Dist

Received 09.09.2025
Received in revised form 17.11.2025
Accepted date 26.11.2025
Published date 30.12.2025

How to Cite: *Herdianto, H., Sihombing, P., Fahmi, F., Tulus, T. (2025). Development of a distance measurement model using a magnification approach and modification of the YOLOv3 architecture. Eastern-European Journal of Enterprise Technologies, 6 (9 (138)), 6–15. https://doi.org/10.15587/1729-4061.2025.339774*

1. Introduction

DAS is a type of robot specifically designed to control cars and replace human drivers. Therefore, capabilities similar to those of a driver must be embedded in DAS, such as detecting objects, measuring distance, turning left and right, stopping, and various other intelligences, with the aim that DAS can make the right decisions for maneuvering or other actions so that collisions with surrounding objects do not occur. Distance is one of the important parameters in DAS that is always monitored for changes when DAS is operating on the road. Therefore, in measuring the distance of an object, issues such as time, accuracy, error, and measurement range become a special concern. This is because if the time, error, accuracy, and measurement range of the object's distance do not meet the established standards, it will have an impact on the accuracy of the decisions made by the DAS, and the possibility of a collision with surrounding objects may occur.

Currently, there are several methods that DAS can use to measure the distance of an object, including light detection and ranging (LIDAR), radar, and cameras. The use of LIDAR to measure the distance of an object is more expensive than cameras [1–3] but has advantages such as good image quality [4]. Meanwhile, radar can measure distances of up to more than 150 meters [5] but has poor image resolution [6].The use of cameras to measure objects can be done with mono cameras and stereo cameras. The use of stereo cameras to measure object distances has the advantage of being able to measure distances of up to 100 meters, but it also has the disadvantages of being more expensive than mono cameras [7], large errors in adjusting images from two cameras, and complicated stereo calibration [2, 6]. Meanwhile, the disadvantages of mono cameras include the need for complex algorithms to measure distance [8]. But cheaper than LIDAR, radar, and stereo cameras [2, 7].

Based on the weaknesses of the three measurement methods above: LIDAR, radar, and stereo cameras, researchers have attempted to improve them. Therefore, research focusing on the development of measurement models, the formulas used, and modifications to the YOLOv3 architecture to improve object distance measurement errors, measurement range, speed, and lower implementation costs has scientific relevance and can be applied practically in the automotive and measurement industries.

2. Literature review and problem statement

To measure the distance of an object using traditional geometry, a monocular and stereo camera can be used. The characteristics of distance measurement using traditional geometry include the use of camera parameters and image structure. Research related to distance measurement using traditional geometry has been conducted extensively, including by [9]. This research was conducted to address problems in soccer robots so that they can measure the distance of objects. This research describes two ways that soccer robots can be used to measure distance, namely by comparing the actual height of the object with the height of the object image and from the location of the object in the image matrix. However, the first method is difficult to apply when the overall shape of the object is not clearly visible and becomes part of the problem in this study that has not been discussed because it requires a more specific additional algorithm. The next study is [10], which was inspired by the significant errors in determining the orientation and angle of the camera used to calculate the distance of objects on the Sony Aibo robot. The distance to the object is measured by calculating the camera's tan φ angle and then multiplying it by the camera's height after subtracting the height of the object. This study still has a problem, namely that it can only measure the distance to objects that are lower than the camera's height. This is because the object distance measurement model and formula used do not have the ability to measure distances where the height of the object exceeds the height of the camera. Next, there is [11] which proposes a new approach to measure the distance of an object using camera parameters, geometry, and road contact points. When an object is not symmetrical with the camera, the distance to the object is calculated in two stages: first, the vertical distance of the object to the camera is calculated, then the result of the first stage is divided by the camera's cos angle to the object. This study leaves unanswered the question of determining the cos angle of the camera to a dynamic object, which is part of the problem not discussed in this study. Research [12] attempts to use complex log mapping (CLM), which is a new method for measuring the distance of objects on surfaces with varying textures. To measure the distance of objects, the focal length of the camera is multiplied by the ratio of the diameter of the first and second measurement objects, and then the result is subtracted from the camera movement distance between the first and second measurements. Although this research can measure distance, it still has problems in terms of the long measurement time. This is due to the measurement model design used. Then [13] in his research attempted to improve the accuracy of object distance measurement using the pinhole approach. In this study, to calculate the distance [13], the camera's focal length variable was multiplied by the width of the object and then divided by the number of pixels in the width of the object. This study has a limitation in that only certain objects can be measured, which is part of the problem in this study that has not been discussed because fundamental improvements are needed in the object detection algorithm used. Research conducted by [14] attempted to measure the distance between objects and vehicles using a camera with low resolution, and the objects detected were light emitting diodes (LEDs) located on the vehicle in front. The parameters used to measure distance included focal length, distance between the two cameras, the angle cosϕ of the camera relative to the LED, and the dimensions of the camera sensor. These parameters were then formulated to obtain the ideal distance formula. Although the results were quite good, there were several problems when applied, such as higher costs due to the use of two cameras and the complexity of camera calibration. Therefore, the use of a low-resolution monocular camera could be a solution. However, this was not discussed in this study due to the complexity of the algorithm that would be used. Next, [15] tries to fix the problem of object distance measurement errors caused by changes in the pitch and roll angles of the vehicle while on the road. The way to find out the magnitude of the changes in these two angles is by measuring inertia, which is then used to update the rotation matrix and correct the vanishing line position to improve distance estimation accuracy. The variables used to measure object distance include camera focal length, camera height, vanishing line value, and the lowest point of the vehicle being measured. The determination of a fixed vanishing line that does not follow changes in the pitch and roll angles of the vehicle is an interesting topic that has not been discussed in this study. In paper [16], the problem of object distance measurement error is corrected by using changes in the camera's viewing angle while driving. The first step is to find the vanishing point based on the texture orientation, followed by calculating the yaw angle and pitch angle of the camera. Then, the distance of the object can be calculated by dividing the height of the camera by the yaw angle and pitch angle of the camera that has been subtracted.

Several studies related to object distance measurement using deep learning have been conducted, such as [17] which attempts to measure object distance using the VGG16 and RES50 architectures that have been modified with objects used by pedestrians, cars, and cyclists obtained from KITTI. The results of this study state that the proposed measurement model can measure object distance and is better than similar distance measurement models such as support vector regressor (SVR) and inverse perspective mapping (IPM), which can be proven from the error results obtained. Next, there is [18] which uses the RES50 architecture by adding a region proposal network (FPN) to the backbone to extract features in many images and geometries. For objects used by pedestrians, cars, and cyclists obtained from KITTI. This study states that the proposed deep learning distance measurement model can measure object distances and is better than similar distance measurement models such as the baseline. This can be seen from the distance measurement error for all objects used, which is lower than the baseline. However, the use of geomnet requires high costs when implemented. Then there is [19] which proposes a modified YOLOv3 architecture by dividing its output into two classes: agnostic and aware. The objects measured are pedestrians, cars, and trucks from KITTI. Both classes can measure object distance, but the agnostic class is better than the aware class because it has a smaller measurement error. Paper [7] proposes the YOLO Light-Fast architecture combined with multiscale prediction YOLOv5 and Shufflenetv2 to improve measurement error and increase object detection speed. The results of the study show an improvement in measurement error compared to previous studies such as geoNet, zhou, struct2depth, monodepth2, and synDistNet. In terms of object detection speed, there is also an improvement compared to previous studies, namely YOLOv5s. Next, [20] proposes a new approach to measuring object distance by modifying convolutional support estimator networks (CSEN). CSEN aims to enable the network to calculate direct mapping for support estimation (SE) tasks in representation-based classification problems. The results of this study indicate that the proposed measurement model can measure object distances and is better than similar distance measurement models such as support vector regressor (SVR) and BaseModel (CRC-light). This can be proven from the smaller error results obtained.

From several studies on traditional geometric distance measurement and deep learning, it is known that each researcher proposes different methods with the aim of improving distance measurement errors, and the results of these studies can all be used to measure distances for different problems. In addition, researchers also analyzed several advantages and disadvantages of traditional geometry and deep learning research. The advantage of traditional geometry is that it uses focal length and object image height variables in measuring distance so that measurement errors can be reduced because they reflect the actual conditions of the object. Meanwhile, the disadvantage is that it uses the height and angle of the camera, which affects its inability to measure the distance of objects that are higher than the camera. In addition, the camera angle is prone to change due to vibrations while driving, which increases measurement errors, so that only certain objects can be measured. Meanwhile, the advantages of deep learning research include faster distance measurement and the ability to measure a larger number of objects. However, its disadvantages include more complex measurement algorithms, and some deep learning methods require significant costs when implemented. Based on the advantages and disadvantages of both studies, as well as the consideration of ease of implementation in the watershed at a low cost, the researcher attempted to add a magnification method to the distance measurement formula for greater precision and used a modified YOLOv3 architecture to measure distance.

3. The aim and objectives of the study

The aim of this study is to develop an object distance measurement model using a monocular camera combined with magnification theory and modifications to the YOLOv3 architecture. The developed object distance measurement model is expected to be faster in the distance measurement process and reduce the error in object distance measurement results.

To achieve this aim, the stages of research carried out are as follows:

– to repeat several measurement models from previous studies to identify weaknesses and validate the results;

– to collect initial distance measurement data for several objects at specified distance intervals and maximum range;

– to understand the parameters of the monocular camera and the theory of magnification in an effort to find the right formula that can be used for object distance measurement;

– to modify the YOLOv3 architecture used to detect objects, measure object image height, and measure distance.

4. Materials and methods

4. 1. Object and hypothesis of the study

The object of the study is measuring distances to objects.

The main hypothesis of this study is that Hybrid Dist – YOLOv3, which is a combination of the magnification approach and the proposed modification of YOLOv3, can be used to measure distance with a measurement time of < 10ms, can measure the distance to objects that are actually taller than the camera, and the ε_A value of the modified YOLOv3 architecture is lower than the results of the study [19].

The assumption made by researchers in this study is that the height of the object image, which is one of the variables in the distance measurement formula, is equal to the height of the bounding box in YOLO. In order to minimize measurement error, the percentage of YOLOv3 object detection results must be ≥ 95%.

Simplifications adopted in the study are that the amount of image data for each object class to be used in training can reach thousands. Some of these object images have similarities in terms of the actual height of the object and the shape of the object. Therefore, to reduce the computational load and training time, the training data can be simplified by not including all image data of objects that have similarities, but only selecting a few.

4. 2. Camera specifications and data

For camera specifications used during initial data collection and used when compiling distance formulas as shown in Table 1 while data from three class samples used as shown in Table 2.

Table 1

Specifications of the camera used

Name	Value
Model camera	Canon A2300
Focal length	35 mm
Image size	3456*4608
Sensor dimensions	11.04 mm

Table 2

Actual height of object

No.	Data object	
	Name object	Actual object height (meters)
1	Jerrycans	0.170
2	Bottle	0.260
3	Person	1.750

Table 2 shows that the three object samples used have different heights. With three samples of different object heights, researchers can conclude that the actual height of the object is one of the variables used in determining the distance formula.

4. 3. Evaluation criteria

To determine how well the YOLOv3 architecture performs in detecting objects, researchers measured it using the mean Average Precision (mAP) metric. This metric is the standard for object detection using the common objects in context (COCO) dataset [8] and was also used in the study [7]. Meanwhile, to determine the performance of the proposed YOLO architecture in measuring distance, researchers used the same measure as [19], namely absolute distance error (MAE) expressed as ε_A and relative average expressed as ε_R

$$\varepsilon_A = \frac{1}{n}\sum_{n=1}^{n}\left|d_i - \hat{d}_i\right|, \tag{1}$$

whereas for ε_R

$$\varepsilon_R = \frac{1}{n}\sum_{n-1}^{n}\frac{\left|d_i - \hat{d}_i\right|}{\max\left(d_i,1\right)}, \quad (2)$$

where d – the actual distance and $\hat{d}$ – the predicted measurement. For n is the number of bounding boxes found.

5. Results of the Hybrid Dist-YOLOv3

5. 1. Reengineering in determining the best model for measuring distance with a monocular camera

To determine the best model to use for measuring object distance with a monocular camera, the researchers redesigned two object distance measurement models as shown in Fig. 1. The aim was to identify the shortcomings of the models used. The object distance measurement model in Fig. 1, *a*, uses the camera height and the size of the camera tangent angle (φ) to measure distance, so it has the disadvantage of not being able to measure the distance of objects if the height of the object being measured is higher than the camera height. Meanwhile, the object distance measurement model in Fig. 1, *b*, uses image height to measure the distance of the object from two images by moving the camera along a specified distance. The disadvantage of this distance measurement model is that it takes longer because of the camera shift during the object distance measurement process. After identifying the shortcomings of the previous object distance measurement model, several factors were considered by researchers in determining the object distance measurement model to be used, such as the measurement time must be < 10 milliseconds (to prevent delays in decision-making when the DAS turns left or right, reduces speed, and other actions affecting the forward movement of the DAS), the measurement range must be >50 meters for certain objects such as trucks (to maintain the ideal distance between vehicles according to speed), not using the height and angle of the camera as shown in Fig. 1, *a*, utilizing the camera's focal length and changes in the height of the object image as shown in Fig. 2. Data related to focal length, camera sensor dimensions, and the actual height of each object class uses the data in Tables 1, 2. Based on the above factors, the distance measurement model is determined as shown in Fig. 2.

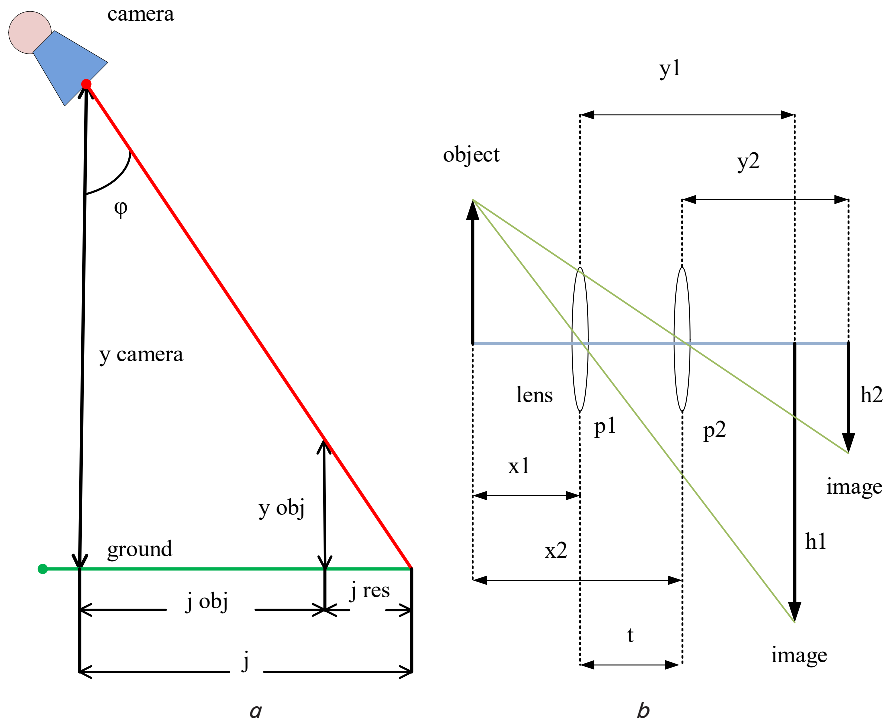

Fig. 1. Object distance measurement model: *a* — distance measurement model [10]; *b* — distance measurement model [12]; y camera — camera height; y obj — object height; j obj — object distance; j — maximum distance; j res — maximum distance minus the object distance; x1, x2 — object distances from the camera; y1, y2 — distances between the lens and the camera sensor; t — camera displacement distance; h1, h2 — object image heights

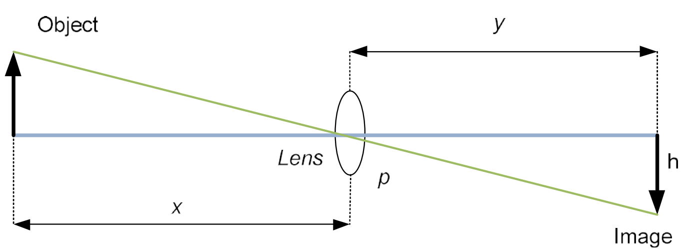

Fig. 2. Proposed distance measurement model: *y* — camera focal length; *h* — height of the object image; *p* — camera lens; *x* — distance of the object from the camera lens

From Fig. 2, the researchers made an initial hypothesis regarding the distance measurement formula to be used. Where the hypothesis is that the focal length and the height of the object image are constants and variables that can be included in the formation of the distance measurement formula. And the distance measurement time will be faster than the object distance measurement model in Fig. 1, *b*, because it does not require camera movement.

5. 2. Collecting initial distance measurement data for several objects

As study material for determining the object distance measurement model, the formula to be used and for training deep learning YOLO and testing data, sample object image data is needed. As initial data, the researcher determined three classes, namely persons, bottles and jerrycans. The blue vertical line in Fig. 3, represents the image height of person object with 1866 pixels at a distance of 3.4 meters. The image height of this object was measured using MATLAB 2019 originated in the United States.

For the person class, 30 samples were taken (only 13 are shown in Table 3), 13 for bottles, and 13 for jerrycans. The details of the object distance measurement samples for each class are shown in Tables 3–5. Distance measurements were taken at 1 meter intervals for person objects, while measurements for bottle and jerrycan objects were taken at 0.3 meter intervals. This change in interval was necessary to determine changes in the image height of the objects.

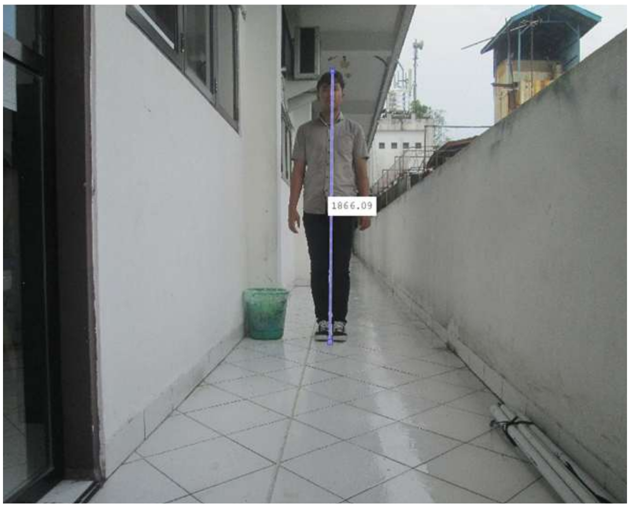

Fig. 3. Measurement of the distance to person object at 3.4 meters: 1866 pixels

Table 3

Measurement data for the height of person image objects

No.	Object name	Object distance (meters)	High object image (pixels)
1	Person	2.4	2526
2	Person	3.4	1866
3	Person	4.4	1440
4	Person	5.4	1182
5	Person	6.4	1008
6	Person	7.4	882
7	Person	8.4	774
8	Person	9.4	700
9	Person	10.4	627
10	Person	11.4	573
11	Person	12.4	525
12	Person	13.4	487
13	Person	14.4	457

Table 4

Measurement data for the height of the bottle object image

No.	Object name	Object distance (meters)	High object image (pixels)
1	Bottle	0.6	1530
2	Bottle	0.9	1038
3	Bottle	1.2	792
4	Bottle	1.5	643
5	Bottle	1.8	545
6	Bottle	2.1	465
7	Bottle	2.4	412
8	Bottle	2.7	369
9	Bottle	3.0	330
10	Bottle	3.3	301
11	Bottle	3.6	274
12	Bottle	3.9	254
13	Bottle	4.2	235

Table 5

Measurement data of the height of the jerrycan object image

No.	Object name	Object distance (meters)	High object image (pixels)
1	Jerrycan	0.6	1015
2	Jerrycan	0.9	692
3	Jerrycan	1.2	513
4	Jerrycan	1.5	426
5	Jerrycan	1.8	352
6	Jerrycan	2.1	300
7	Jerrycan	2.4	265
8	Jerrycan	2.7	236
9	Jerrycan	3.0	212
10	Jerrycan	3.3	194
11	Jerrycan	3.6	178
12	Jerrycan	3.9	164
13	Jerrycan	4.2	154

5. 3. Development of magnification formulas for measuring object distances

Based on the sample data of object distance measurements obtained in Tables 3–5, it is known that the image height of an object will be smaller when it is farther from the camera, and the change is linear. Therefore, a distance formula was developed by finding the relationship between object distance, focal length, sensor dimensions, image height of the object, and actual height of the object.

Magnification is the enlargement which is the ratio between the image height of the object and the actual height of the object. Expressed as

$$\text{magnification}\,(\text{m}) = \frac{\text{high image of the object}}{\text{actual image of the object}}. \quad (3)$$

Meanwhile, field of view (FoV) is the area outside the camera that can be captured, so the size of the FoV is greatly influenced by the dimensions of the sensor used. Therefore, the relationship between magnification and FoV can be written as

$$m \propto \frac{1}{\text{FoV}},$$

next, to obtain the value m that is equal to FoV, it is multiplied by the sensor dimensions to become:

$$m = \text{sensor dimension} * \frac{1}{\text{FoV}}, \quad (4)$$

$$\text{FoV} = \frac{\text{sensor dimension}}{m}. \quad (5)$$

When the focal length is inserted into (5), a new equation is obtained

$$\text{Distance object} = \frac{\text{focal lenght} * \text{sensor dimension}}{m}, \quad (6)$$

and if the value of m is changed by (3) so that it becomes

$$\text{Distance object} = \frac{\text{focal lenght} * \text{sensor dimension}}{\dfrac{\text{high image of the object}}{\text{actual image of the object}}}. \quad (7)$$

This formula will later be entered into YOLO.

5. 4. Modification of the YOLOv3 architecture

The original YOLOv3 architecture is shown in Fig. 4, specifically in the red box. This architecture can only detect objects; it cannot measure the distance of objects.

Therefore, in order to perform object distance detection and measurement tasks, modifications need to be made to the YOLOv3 architecture, as shown in Fig. 4, specifically in the box outlined in green. The workflow for modifying the YOLOv3 architecture is as follows:

Check the percentage of predicted results if ≥ 95 is used to determine the percentage of object detection results. This is done by determining the Intersection over Union (IoU), which is a calculation between two bounding boxes (object bounding box with anchor bounding box). If the object detection prediction result is greater than or equal to 95%, the process will proceed to the select array class actual object height stage, which contains the actual height of each object class as shown in Table 6.

Table 6

Contents of the object class array

Grade	Class name	Actual height of the object (mm)
0.1	Pedestrian	1700
0.2	Car	2000
0.3	Truck	3000

Table 6 is an example of an object class array with class values output from YOLO as follows: 0.1 for the pedestrian class contains the actual object height of 1700; next, 0.2 for the car class contains the actual object height of 2000; 0.3 for the truck class contains the actual object height of 3000. The next process is to get the actual height of the object in the array (*O*), which is the actual height of the object stored as variable *O*. Then the next step is to calculate the height of the bounding box as the height of the object image (*T*), which is calculating the height of the bounding box that is assumed to be the height of the object image. Fig. 5, illustrates the dimensions of the bounding box formed from the coordinates xmin, ymin, xmax, and ymax.

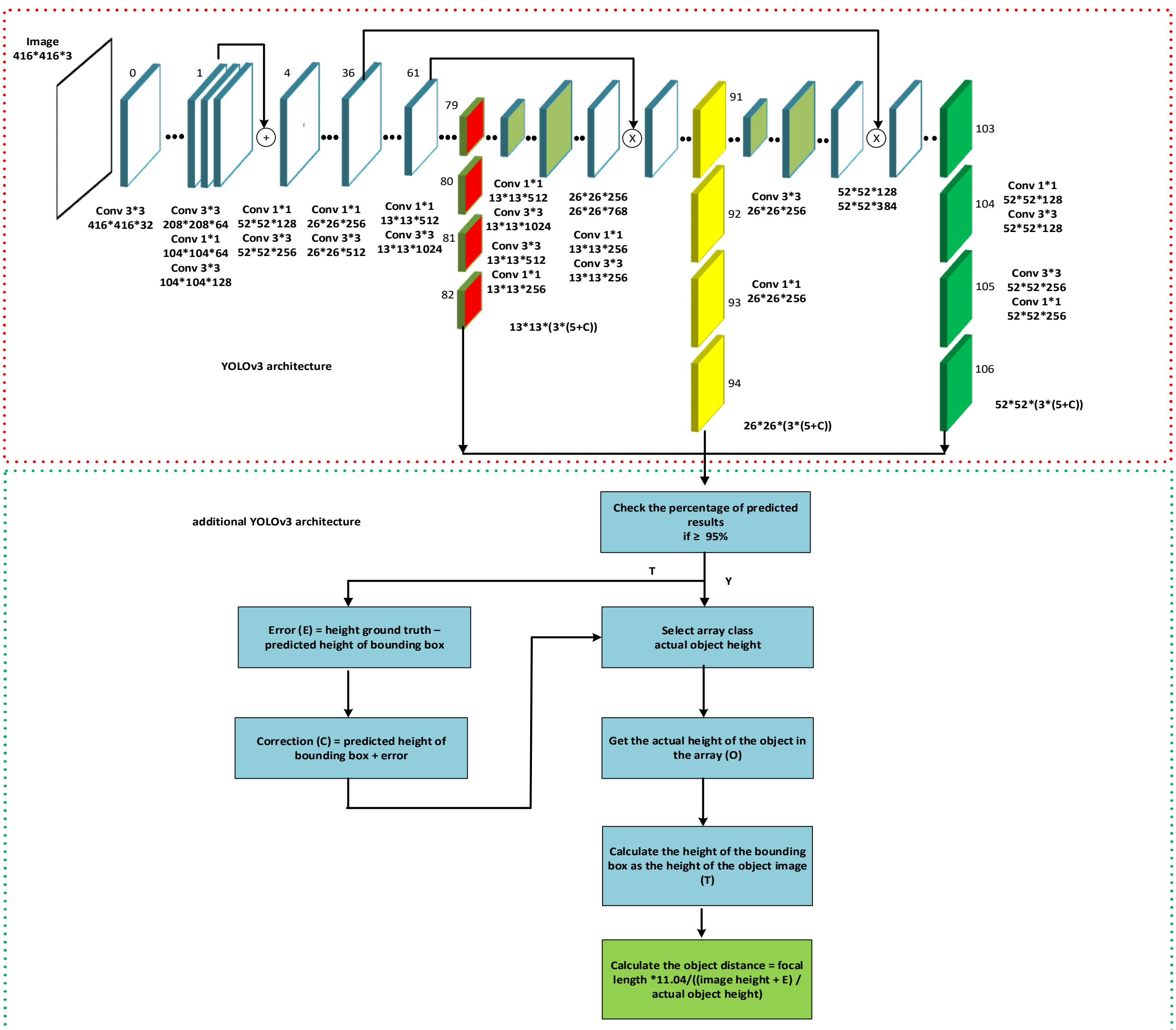

Fig. 4. Modification of the YOLOv3 architecture

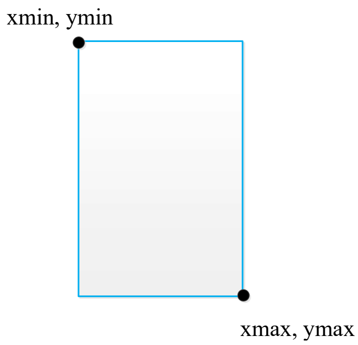

Fig. 5. Coordinate points forming the bounding box

Xmin represents the starting row and xmax represents the ending row, which is then used to represent the height of the object image. To obtain the height of the object image, subtract xmax – xmin in pixels. Meanwhile, ymin represents the starting column and ymax represents the ending column, so that ymin and ymax can be used to represent the width of the object image. After the values of the variables O and T are known, proceed to the stage of calculating the object distance. The distance to the object is calculated using (7) plus E (distance = focal length *11.04 / ((image height + E) / actual object height) with the assumption that the value of variable $E = 0$ because object detection is already ≥95%. This means that the results of the object distance measurement will follow the detection results and this value is already as expected. For some objects that have never been recognized (not trained) in the YOLO network, sometimes when tested, the object detection is < 80%. This result is not as desired, so to overcome this, an error calculation is performed (E) = ground truth height – predicted height of bounding box addition of error correction.

In order for the distance measurement model to detect and measure the distance of objects, the modified YOLO architecture needs to be trained on the objects to be detected, such as people, bottles, and jerrycans at various distances. Of the 56 image data collected, 80% was used for training and 20% for testing. After the training results were deemed satisfactory based on the number of epochs or error tolerance achieved, testing was continued with the training and testing data as shown in Table 7.

Table 7

Results of detection and measurement of object distance with initial data

Class	Number	$mAP_{0.50}$	Distance (meters)			
			Min	Max	ε_A	ε_R
Person	30	0.994	–0.121	1.020	–0.274	0.010
Bottle	13	1	–0.097	0.140	–0.153	–0.066
Jerrycans	13	1	–0.099	0.226	–0.163	–0.086

For ε_R, the relative error is 100% when the value is 1.

From Table 7, it can be seen that the IoU value used is 50% for all objects, where the test results show that the AP values for persons, bottles, and jerrycans are 0.994, 1, and 1. Meanwhile, for the distance of person, bottle, and jerrycan objects, the values are ε_A –0.274, –0.153, –0.163. These absolute error values indicate that the measurement error tolerance for object distance is quite low and in line with the desired value of < 0.5 meters. Fig. 6, shows the test of measuring the distance of a person object at 5.4 meters.

Fig. 6. Results of YOLOv3 distance measurement predictions

YOLOv3 can correctly recognize the object (indicating a person) with a measured distance of 563.969 cm or 5.639 meters and a measured bounding box height of 1199 pixels.

After the performance of the modified YOLO network was deemed satisfactory, the next step was to retrain the modified YOLO network with the dataset from the Karlsruhe Institute of Technology and Toyota Technological Institute (KITTI) because the number of objects and class names were different from the dataset used in the preliminary testing. This was done to compare the results of the previous object distance measurements using different methods with the proposed method. The reasons for using the KITTI dataset include the fact that previous research [19] used this dataset, and the variables needed to form the proposed object distance measurement formula (actual object height) and the actual object distance used as a reference for measuring distance error are in the KITTI dataset. Meanwhile, the focal length and camera sensor dimensions used are not explicitly stated, so researchers must look for them in the camera datasheet used. In this study, the focal length of the camera and the dimensions of the sensor used were set at 12 and 5.84, respectively. Only three classes were used in the training, namely pedestrian, car, and truck. Furthermore, to determine the performance of the modified YOLO network, tests were conducted as shown in Fig. 7–9.

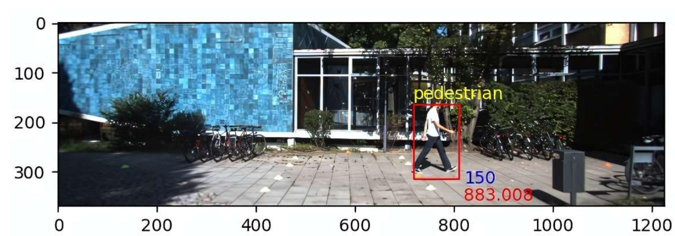

Fig. 7. Measurement of pedestrian object distance at a distance of 8.4 meters

Fig. 7 shows the measurement test of the pedestrian object at an actual distance of 8.4 meters. The value 150 represents the height of the object image in pixels, while 883.008 is the distance produced by YOLO in cm or 8.83008 meters.

Fig. 8 shows the distance measurement test for the two detected cars. Car 1 has an actual distance of 12.58 with a measured distance of 1245.866 cm or 12.45 meters with an object image height of 90 pixels. Car 2 has a distance of 24.19 with a measured distance of 2076.44 cm or 20.764 meters with an object image height of 54 pixels.

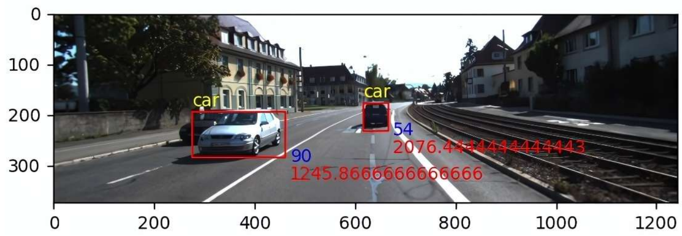

Fig. 8. Measurement of the distance of the object car at distances of 12.58 and 24.19 meters

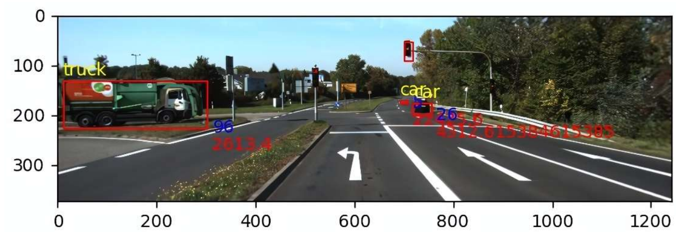

Fig. 9. Measurement of the distance of the truck at a distance of 28.2 meters

Fig. 9 shows the measurement test of the truck object at an actual distance of 28.20 meters. The value 96 represents the height of the object image in pixels, while 2613.4 is the distance produced by YOLO in cm or 26.134 meters. Overall, the distance measurement error values for each class can be seen in Table 8.

Table 8

Comparison of ε_A and ε_R values from previous research measurements with those proposed

Class	Number	Dist-YOLOv3G		Dist-YOLOv3W		Hybrid Dist YOLOv3	
		ε_A	ε_R	ε_A	ε_R	ε_A	ε_R
Pedestrian	1150	1.75	0.11	4.87	0.41	1.37	0.08
Car	7682	2.49	0.11	10.68	0.44	2.15	0.09
Truck	318	4.63	0.14	6.49	0.21	3.74	0.11

From Table 8, it is known that there are three object distance measurement methods compared, namely Dist-YOLOv3G, Dist-YOLOv3W, and Hybrid Dist-YOLOv3. Two methods are from previous research (Dist-YOLOv3G, Dist-YOLOv3W) and one method (Hybrid Dist-YOLOv3) is proposed. Using the same classes and number of objects as in the previous study, object distance measurement testing was conducted, with the size used as an indicator of improvement ε_A, ε_R. The smaller the values of ε_A, ε_R obtained from the method used, the better the method is compared to the others.

6. Discussion of the Hybrid Dist-YOLOv3 results

Hybrid Dist is a distance measurement that integrates traditional object distance measurement and deep learning. For traditional distance measurement, this method is based on a formula that utilizes camera parameters and image structure. Meanwhile, deep learning measurement relies on the use of the YOLOv3 architecture to detect and measure the height of objects in images. Fig. 2 shows the object distance measurement model proposed in this study. The proposed model differs from previous distance measurement models, such as those used in [9–12, 15], and is an improvement on the distance measurement model used in [12]. The proposed distance measurement model has advantages over similar object distance measurement models, such as only using the focal length, camera sensor dimensions, and image height. This model does not use camera height and angle, so object distance can still be measured even if the object height is higher than the camera height. In addition, the time required to measure object distance is faster when using this model compared to CLM [12]. The reason for the above statement is that the proposed object distance measurement model only requires three steps, while CLM requires nine steps to complete object distance measurement.

To obtain the object distance measurement formula in accordance with the proposed distance measurement model, the first step taken by the researcher was to take samples of distance measurements of several classes of objects at various predetermined distances. Then each class sample was measured for object image height as shown in Fig. 3 using the MATLAB 2019 tool. The results of the object distance measurement samples are shown in Tables 3–5. The object distance measurement sample data obtained was then analyzed, and it was concluded that distance can be measured using a monocular camera because the image height changes linearly based on changes in object distance. Then, referring to the formula used in previous studies [12, 15] several constants and variables were determined to be used in the formulation of the object distance measurement formula. These constants are focal length and camera sensor dimensions, while the variable is the height of the object image. Next, a substitution experiment was conducted on the constants and variables used to measure the distance of an object in order to obtain an accurate formula for measuring the distance of an object according to the distance measurement model used, which could then be incorporated into the YOLOv3 architecture. Up to this stage, the distance measurement formula results were calculated manually and then compared with the actual distance, and the results always had a large error (did not match the actual distance) when tested on different objects. Finally, an attempt was made to add a variable to the distance measurement formula, namely by entering the actual height of the object into the distance measurement formula. This addition of a variable is known as magnification, and after being tested with sample data from three objects in Tables 3–5, the measurement results were declared to be very good.

Generally, YOLOv3 is used for classification and object detection. To be able to measure object distances, modifications need to be made to its architecture as shown in Fig. 4. However, the architectural modifications to YOLOv3 do not necessarily have to be as shown in Fig. 4. They can take other forms. The modified YOLOv3 architecture must still undergo a training process with the initial data samples in Tables 3–5 to obtain the ideal network weights. Once the training results are deemed satisfactory, testing with the initial data samples can proceed. The first objective of testing at this stage is to prove the initial hypothesis that the height of an object image can be measured using YOLOv3 by counting the number of pixels in the bounding box of the formed object. The results displayed by YOLOv3 are very close to the manual measurement results of the object image height using MATLAB 2019. The second objective is to determine the accuracy and error rate of object distance measurements based on the initial data sample. Table 7 shows the results obtained from YOLOv3 in measuring the distance of objects with mAP 0.50 on human, bottle, and jerry can objects, which are 0.994, 1, and 1, respectively, while the $\boldsymbol{\varepsilon}_A$ values are –0.274, –0.153, and –0.163. The testing continued by comparing the research results with other previous research methods [19] with the object classes tested, namely pedestrians, cars, and trucks from the KITTI dataset. The results are shown in Table 8, where the $\boldsymbol{\varepsilon}_A$ values for the three object classes show an improvement from the previous study, which were originally 1.75, 2.49, and 4.63, to 1.37, 2.25, and 3.74.

If the results of this study are applied in the field, there are several limitations in its application, such as the use of cameras that differ from the specifications of the cameras used in this study, then the focal length and sensor dimensions used in the magnification formula (7) must be changed according to the focal length and sensor dimensions of the camera used. Next, the actual height of the object which distance is to be measured must be known in advance, and then the height of the object is entered into the array class. If these limitations are not addressed, the results of the object distance measurement will have a significant error. Although the results of this study are better than those of [19], according to the researchers, there are still weaknesses, such as some objects with an actual distance of > 50 meters, causing the image size of the object to be smaller than 20 * 20 pixels, making it difficult to detect by the YOLOv3 architecture used in this study. For further research related to overcoming the weaknesses of this study, the researcher suggests three improvements, namely improvements to the main architecture of YOLOv3, the development of a multi-level magnification algorithm, and combining the two previous improvements.

7. Conclusion

1. From the redesign of two previous object distance measurement models, it is known that the first distance measurement model uses the camera height and the camera tangent angle (φ) to measure the distance of the object. The second object distance measurement model uses the object image height to measure the distance of the object from two object images obtained by moving the camera along a certain distance. Meanwhile, the proposed object distance measurement model uses the camera focal length and the object image height.

2. From the initial data on object distance measurements at various intervals, it is known that there is a linear change in the height of the object image relative to the distance measured based on the number of pixels. Due to the change in the height of the object image, it can be concluded that the object distance can be measured where the height of the object image is used as one of the variables in the object distance measurement formula used.

3. The object distance measurement formula used is obtained from the relationship between field of view, focal length, sensor dimensions, magnification, and adjustment to the proposed distance measurement model. Manual distance calculations using the proposed formula produce a small error of < 20 cm if the measured image height of the object and the actual height of the object are accurate.

4. Modification of the YOLOv3 architecture in the output section is intended to enable the measurement of object distances. The distance measurements obtained are better than those in previous studies. This is evidenced by the ε_A value obtained, which is smaller than the test results for three sample objects: pedestrian, car, and truck.

Conflict of interest

The authors declare that they have no conflict of interest in relation to this study, whether financial, personal, authorship or otherwise, that could affect the study and its results presented in this paper.

Financing

The study was performed without financial support.

Data availability

Manuscript has data included as electronic supplementary material.

Use of artificial intelligence

The authors confirm that they did not use artificial intelligence technologies when creating the current work.

References

1. Raj, T., Hashim, F. H., Huddin, A. B., Ibrahim, M. F., Hussain, A. (2020). A Survey on LiDAR Scanning Mechanisms. Electronics, 9 (5), 741. https://doi.org/10.3390/electronics9050741
2. Ding, M., Zhang, Z., Jiang, X., Cao, Y. (2020). Vision-Based Distance Measurement in Advanced Driving Assistance Systems. Applied Sciences, 10 (20), 7276. https://doi.org/10.3390/app10207276
3. Kim, J. B. (2019). Efficient Vehicle Detection and Distance Estimation Based on Aggregated Channel Features and Inverse Perspective Mapping from a Single Camera. Symmetry, 11 (10), 1205. https://doi.org/10.3390/sym11101205
4. Haseeb, M. A., Guan, J., Ristić-Durrant, D., Graser, A. (2018). DisNet: A Novel Method for Distance Estimation from Monocular Camera. In 10th Planning, Perception and Navigation for Intelligent Vehicles. Madrid. Available at: https://project.inria.fr/ppniv18/files/2018/10/paper22.pdf

5. Lee, S., Han, K., Park, S., Yang, X. (2022). Vehicle Distance Estimation from a Monocular Camera for Advanced Driver Assistance Systems. Symmetry, 14 (12), 2657. https://doi.org/10.3390/sym14122657
6. Ali, A., Hassan, A., Ali, A. R., Ullah Khan, H., Kazmi, W., Zaheer, A. (2020). Real-time vehicle distance estimation using single view geometry. 2020 IEEE Winter Conference on Applications of Computer Vision (WACV), 1100–1109. https://doi.org/10.1109/wacv45572.2020.9093634
7. Liang, H., Ma, Z., Zhang, Q. (2022). Self-Supervised Object Distance Estimation Using a Monocular Camera. Sensors, 22 (8), 2936. https://doi.org/10.3390/s22082936
8. Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D. et al. (2014). Microsoft COCO: Common Objects in Context. Computer Vision – ECCV 2014, 740–755. https://doi.org/10.1007/978-3-319-10602-1_48
9. Jamzad, M., Foroughnassiraei, A., Chiniforooshan, E., Ghorbani, R., Kazemi, M., Chitsaz, H. et al. (2000). Middle Sized Soccer Robots: ARVAND. RoboCup-99: Robot Soccer World Cup III, 61–73. https://doi.org/10.1007/3-540-45327-x_4
10. Jüngel, M., Mellmann, H., Spranger, M. (2008). Improving Vision-Based Distance Measurements Using Reference Objects. RoboCup 2007: Robot Soccer World Cup XI, 89–100. https://doi.org/10.1007/978-3-540-68847-1_8
11. Joglekar, A., Joshi, D., Khemani, R., Nair, S., nd Sahare, S. (2011). Depth Estimation Using Monocular Camera. International Journal of Computer Science and Information Technologies, 2 (4), 1758–1763. Available at: https://www.ijcsit.com/docs/Volume 2/vol2issue4/ijcsit2011020480.pdf
12. Venu Gopala Krishnan, J., Manoharan, N., Sheela Rani, B. (2010). Estimation of Distance To Texture Surface Using Complex Log Mapping. J. Comput. Appl., 3 (3), 16–21. Available at: https://d1wqtxts1xzle7.cloudfront.net/30862784/50116157_vol3i3p4-libre.pdf?1392203742=&response-content-disposition=inline%3B+filename%3DEstimation_Of_Distance_to_Texture_Surfac.pdf&Expires=1764668791&Signature=HuHi4o78VYI164-U6Swdr~e5-p6cYWI92CSjmgVeRncJDw8ozK8aHhi5gpCo2sgdVRTsUqMdDArIQmvy2QnPDWsKKriFM7lR4TDt056Sb7~ZCP3hv-YDwZFxT~0nC8fjSOGZhkoVx6NJe0HYeuh87m5VSwtOSb0i0nlrx3N72n4CtoiUXFkKo1Wr4kDaYvi4H~CnwlRZRxHbLKY82fPcIKwAGX9M0oKH76LmuhHw~h1ZEoh2~ntoH6PR9-nRv3dfJu0jIxPNip-q6BZsceOBE0Ksyl4FPjw9tPL4NsIA2kJRv2rA-ecSM2TxHN1xzF19uHoe0newY30eGu~QlT8kBQ__&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA
13. Megalingam, R. K., Shriram, V., Likhith, B., Rajesh, G., Ghanta, S. (2016). Monocular distance estimation using pinhole camera approximation to avoid vehicle crash and back-over accidents. 2016 10th International Conference on Intelligent Systems and Control (ISCO). https://doi.org/10.1109/isco.2016.7727017
14. Tram, V. T. B., Yoo, M. (2018). Vehicle-to-Vehicle Distance Estimation Using a Low-Resolution Camera Based on Visible Light Communications. IEEE Access, 6, 4521–4527. https://doi.org/10.1109/access.2018.2793306
15. Qi, S. H., Li, J., Sun, Z. P., Zhang, J. T., Sun, Y. (2019). Distance Estimation of Monocular Based on Vehicle Pose Information. Journal of Physics: Conference Series, 1168, 032040. https://doi.org/10.1088/1742-6596/1168/3/032040
16. Liu, J., Zhang, R., Hou, S. (2020). Inter-vehicle distance estimation considering camera attitude angles based on monocular vision. Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering, 235 (2-3), 894–900. https://doi.org/10.1177/0954407020941399
17. Zhu, J., Fang, Y. (2019). Learning Object-Specific Distance From a Monocular Image. 2019 IEEE/CVF International Conference on Computer Vision (ICCV), 3838–3847. https://doi.org/10.1109/iccv.2019.00394
18. Zhang, Y., Ding, L., Li, Y., Lin, W., Zhao, M., Yu, X., Zhan, Y. (2021). A regional distance regression network for monocular object distance estimation. Journal of Visual Communication and Image Representation, 79, 103224. https://doi.org/10.1016/j.jvcir.2021.103224
19. Vajgl, M., Hurtik, P., Nejezchleba, T. (2022). Dist-YOLO: Fast Object Detection with Distance Estimation. Applied Sciences, 12 (3), 1354. https://doi.org/10.3390/app12031354
20. Ahishali, M., Yamac, M., Kiranyaz, S., Gabbouj, M. (2023). Representation based regression for object distance estimation. Neural Networks, 158, 15–29. https://doi.org/10.1016/j.neunet.2022.11.011