Effect of Distance and Light Intensity on Multiple Detection Object

Yo Ceng Giap
Faculty of Sains and Technology,
Universitas Buddhi Dharma
Banten, Indonesia
cenggiap@ubd.ac.id

Muljono
Faculty of Computer Science,
Universitas Dian Nuswantoro
Semarang, Indonesia
muljono@dsn.dinus.ac.id

Mohammad Arief Soeleman
Faculty of Computer Science,
Universitas Dian Nuswantoro
Semarang, Indonesia
m.arief.soeleman@dsn.dinus.ac.id

Affandy
Faculty of Computer Science,
Universitas Dian Nuswantoro
Semarang, Indonesia
affandy@dsn.dinus.ac.id

Ruri Suko Basuki
Faculty of Computer Science,
Universitas Dian Nuswantoro
Semarang, Indonesia
ruri.basuki@dsn.dinus.ac.id

2023 6th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI) | 979-8-3503-5834-6/23/$31.00 ©2023 IEEE | DOI: 10.1109/ISRITI60336.2023.10467935

***Abstract*— Technological advances in the field of computer vision have begun to be developed, one of which is object detection. Object detection serves to find out information related to the detected object. Many models can be used to detect objects such as Yolo, Resnet and SSD. In object detection, the problems that arise are the distance between the object and the camera and the light intensity. The lower the light intensity, the lower the camera specifications and the further the detection distance will affect the object detection process. This research aims to analyze the detection accuracy and computation under different distance settings and lighting conditions. Based on the experimental results by comparing three models namely Yolo, Resnet and SSD, it can be seen that the best accuracy rate is SSD with an accuracy rate of 98.19%, the decrease in accuracy rate based on the effect of the distance of shooting with the camera occurs in all models while the light intensity that has the best accuracy value is at a low light intensity of 270 Lux. The findings of this study can contribute to improving object detection systems that rely on SSD for applications such as face detection, autonomous driving, surveillance, and traffic monitoring.**

Keywords*— *Effect, detection object, distance, light intensity, SSD

I. Introduction

Recently, the concept of object detection has been widely discussed by experts. Object detection is one of the most important jobs in computer vision and robotics [1]. In the area of artificial intelligence, object detection is a very important technology [2]. Object detection is one of the developments of the image classification task. The basic idea behind object recognition is to look at the whole picture and decide if a part is an object or part of the background.

Object detection differs from image classification in that it can detect multiple objects in a single image and provide information about each of them. Object detection has many wide-ranging applications, such as in face recognition [3], security surveillance [4], autonomous vehicles [5], augmented reality [6], agricultural [7], sea surface [8] and more.

With the advancement in computer vision and deep learning, object detection has been developed and become one of the important technologies in image analysis and pattern recognition. Methods that can be used for object detection include RCNN [9], Yolo [10], Resnet [11], SVM [12] and Single Shot Multibox Detector [13].

To be able to measure the detection distance of an object can be done in two ways, namely active and passive measurements. Active measurement is done by measuring the distance of the object by sending several signals to the object while passive measurement is done without emitting signals but by using light intensity information. [14].

In doing object detection, the things we need to pay attention to are distance and light intensity. The lower the light intensity, the lower the camera specifications and the further the detection distance will affect the object detection process. [15].

This object detection research focuses on people using three models, namely Yolo, Resnet50 and SSD. The three models were compared by setting the distance and light intensity. The best accuracy level results as a result for the next research.

II. Review of Related Literature

A. You Only Look Once (Yolo)

YOLO is designed for rapid object detection, developed by Joseph Redmon, and first introduced in 2015. YOLO can perform object detection operations in real time. YOLO divides the picture into grids using a single convolutional neural network. Each grid makes predictions about the bounding box and accuracy rate. Then, the best accuracy rate is used to decide the object class [16].

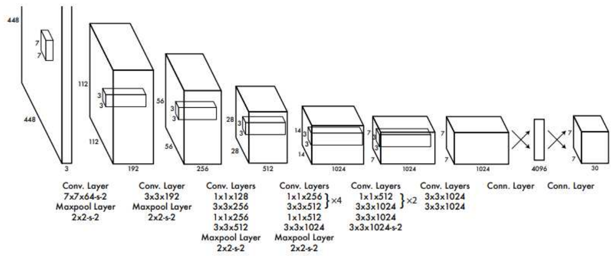

Fig 1. YOLO Architecture [17]

GoogLeNet as an image classifier is the inspiration for the YOLO network architecture. It has twenty-four layers of convolution and then two levels that are fully connected. YOLO uses a 1x1 floating layer, then a 3x3 convolutional layer (See Fig 1).

B. Resnet50

ResNet50 stands for Residual Network and consists of 50 layers. Resnet is a type of neural network architecture that has revolutionized deep learning by enabling much deeper network training with better accuracy than ever before. In 2015 Resnet was first introduced by Microsoft researchers Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. The ResNet50 architecture (See Fig 2) consists of a diverse set of identical layers and ensures that blocks are used to signify the use of previous layers in the network [18].

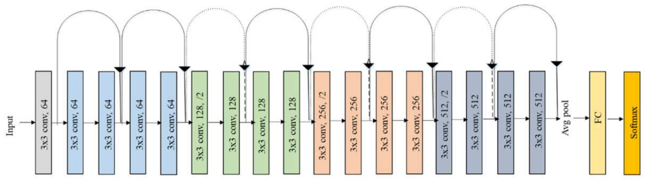

Fig 2. ResNet Architecture [19]

Resnet is used a lot in computer vision jobs like finding objects, classifying images, and segmentation. Resnet has gotten the best results possible on many datasets, such as ImageNet, COCO, and Pascal VOC.

C. SSD

Single Shot MultiBox Detector is a deep learning model used to find items in images or videos. Single Shot Detector is a simple way to solve the problem, but so far it has worked very well [13].

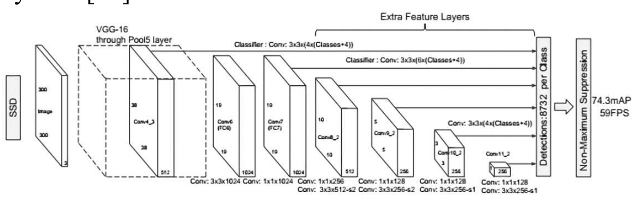

Fig 3. SSD Architecture [13]

In Fig 3, SSD has two network architectures: the VGG-16 network and the network on extra feature layers. The VGG-16 network is used as the base network because it works well for high-quality pictures. Also, SSD adds to the end of the base network a convolutional feature layer that suggests different aspect ratios. Inception, Mobilenet, or Resnet design can be used for the network in the extra feature layer [13].

SSD is a Convolutional Neural Network (CNN) that can be divided into three parts:

1. Base convolutions: a low-level feature map will be produced using convolution taken from an existing image classification architecture.
2. Auxiliary convolutions: the underlying network will be enhanced with convolution to provide a higher level feature map.
3. Prediction convolutions: locating and identifying objects in this feature map by convolution.

D. Intersection Over Union (IOU)

IOU (Intersection over Union) is a phrase used to indicate how much two squares overlap. IOU, sometimes referred to as the Jaccard index (Jaccard similarity coefficient), has been widely used to assess how similar a limited number of samples are to one another [1]. The larger the overlap area, the larger the IOUs.

IOU is primarily utilised in object identification applications, where we train the model to produce a box that properly fits the object. IOU is also utilised in non-maximal suppression, which aims to eliminate boxes that are close to an item based on which box has the highest confidence. In Fig 4, Let's assume that box 1 is represented by [x1, y1, x2, y2], and box 2 is represented by [x3, y3, x4, y4].

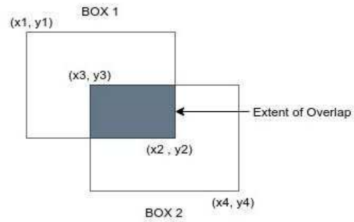

Fig 4. Intersection of two boxes

E. Bounding Box Prediction

To be able to predict an object in the image, the image must be given a bounding box on the object to be predicted. to predict using 4 coordinates for each bounding box, t_x, t_y, t_w, t_h. If the cell is offset from the top left corner of the image by (c_x; c_y), the previous bounding box has width and height p_w, p_h, then prediction can be done using the sigmoid function [20]:

$$b_x = \sigma(t_x) + C_x$$
$$b_y = \sigma(t_y) + C_y$$
$$b_w = P_w e^{tw}$$
$$b_h = p_h e^{th}$$

Where b_x and b_y are the coordinates of the box center, b_w and b_h are the box width and height, c_x and c_y are the filter application locations and t_i is predicted during regression (See Fig 5).

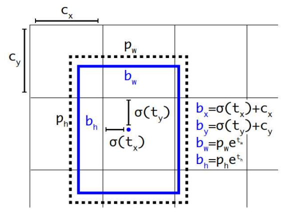

Fig 5. Bounding boxes with dimension priors and location [20]

III. Methodology

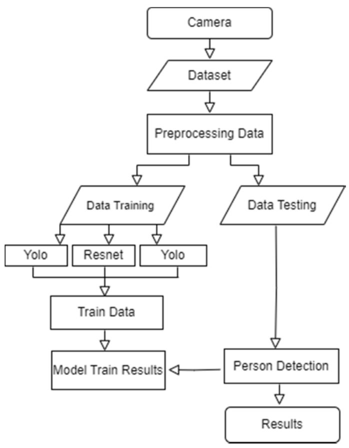

Fig 6. Object detection framework

In Fig 6, The object detection process can use a camera (handphone or webcam), the dataset used consists of public data and private data, the resulting dataset is processed divided into training data and testing data, the training data is processed with the Yolo, Resnet and SSD models, the training results become the model used to test the testing data and the test results are accuracy.

IV. Result And Discussion

In this research, the author conducted several experiments with detection objects focused on people. The datasets used are public datasets using COCO 2017 and private datasets. The private dataset collects its own data by taking video data using a Samsung Galaxy A21s mobile device with video specifications, namely camera: 48 Megapixel, f/2.0; Frame rate: 30fps (1080p). The video was taken in a classroom by following the standardization of light intensity issued by the Badan Standardisasi Nasional (BSN) regarding energy in lighting systems [21].

In the experiment conducted by the researcher, the researcher used a Lux Meter measuring instrument to determine the two lighting categories taken, namely normal light with a light intensity of 362 Lux and a light intensity of less than 270 Lux. The resulting video is 5 seconds long with a frame rate of 30 f/s, frame width 1600 and frame height 720. The following is the video data used:

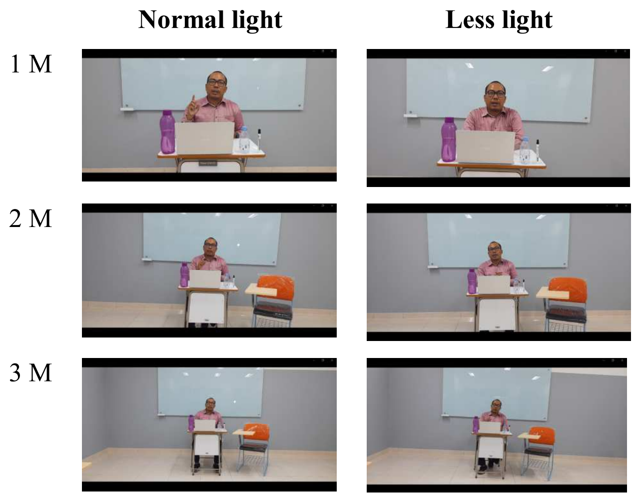

Fig 7. Video Data

From the video that has been made, then the video file is converted from video to image using the VLC media player application. The following is the number of images generated for each distance and light intensity:

TABLE1. Total Images

DISTANCE	LIGHT	NUMBER OF PICTURES
1 Meter	Normal Light	127
1 Meter	Less Light	128
2 Meter	Normal Light	129
2 Meter	Less Light	127
3 Meter	Normal Light	127
3 Meter	Less Light	122
TOTAL PICTURES		**760**

From a total picture of 760, a model test will be carried out using 3 models, so that the image results after the test process are 760 x 3 = 2.280 images.

A. *1 Meter distance and normal light*

The first experiment was conducted indoors with an object distance of 1 meter and normal lighting so as to get the accuracy level of each model with the same object, the results obtained are as follows:

TABLE 2. Accuracy results for a distance of 1 meter and normal light

	YOLO	RESNET	SSD
Min	88.86%	94.09%	30.71%
Max	99.97%	97.11%	99.97%
Average	99.76%	96.01%	98.82%

The figure below shows the findings of the comparison of the three models' degrees of accuracy:

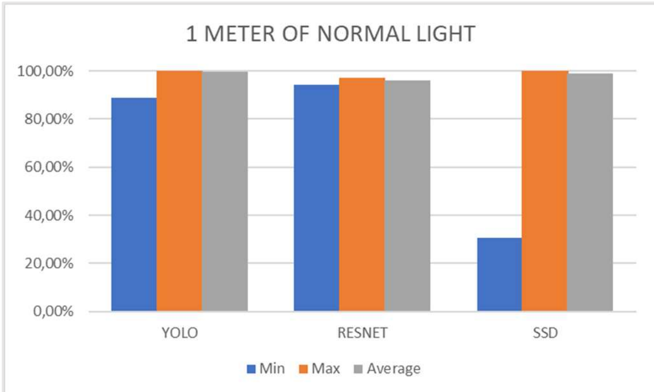

Fig 8. Results graph for 1 meter and normal light

From the data results above, it can be seen that the best accuracy rate for a distance of 1 meter with normal light is Yolo with an accuracy rate of 99.76%.

B. *1 Meter distance and less light*

The second experiment was conducted indoors with an object distance of 1 meter and less lighting so as to get the accuracy level of each model with the same object, the results obtained are as follows:

TABLE 3. Accuracy results for a distance of 1 meter and less light

	YOLO	RESNET	SSD
Min	99.99%	95.12%	98.87%
Max	100.00%	97.51%	99.98%
Average	100.00%	96.14%	99.85%

The figure below shows the findings of the comparison of the three models' degrees of accuracy:

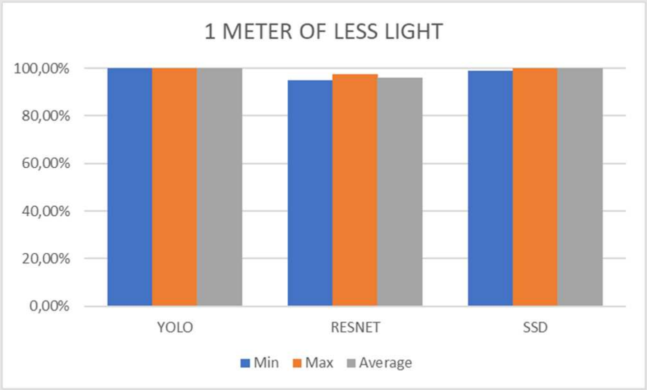

Fig 9. Results graph for 1 meter and less light

From the data results above, it can be seen that the best accuracy rate for a distance of 1 meter with less light is Yolo with an accuracy rate of 100%.

C. 2 Meter distance and normal light

The third experiment was conducted indoors with an object distance of 2 meters and normal lighting so as to get the accuracy level of each model with the same object, the results obtained are as follows:

TABLE 4. Accuracy results for a distance of 2 meter and normal light

	YOLO	RESNET	SSD
Min	98.59%	81.27%	98.96%
Max	99.97%	88.44%	99.90%
Average	99.65%	85.02%	99.70%

The figure below shows the findings of the comparison of the three models' degrees of accuracy:

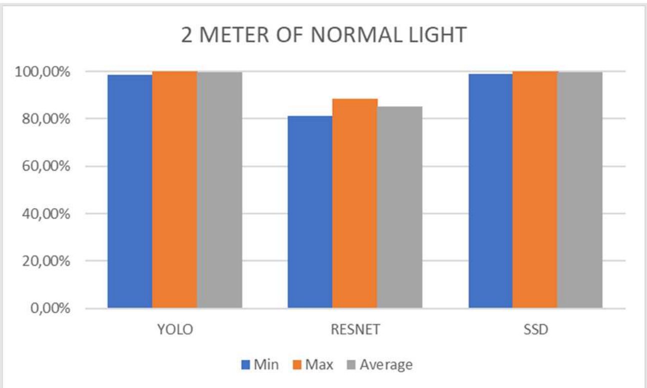

Fig 10. Results graph for 2 meter and normal light

From the data results above, it can be seen that the best accuracy rate for a distance of 2 meters with normal light is SSD with an accuracy rate of 99.70%.

D. 2 Meter distance and less light

The fourth experiment was conducted indoors with an object distance of 2 meters and less lighting so as to get the accuracy level of each model with the same object, the results obtained are as follows:

TABLE 5. Accuracy results for a distance of 2 meter and less light

	YOLO	RESNET	SSD
Min	98.97%	84.39%	91.77%
Max	99.99%	92.92%	99.95%
Average	99.88%	88.55%	99.78%

The figure below shows the findings of the comparison of the three models' degrees of accuracy:

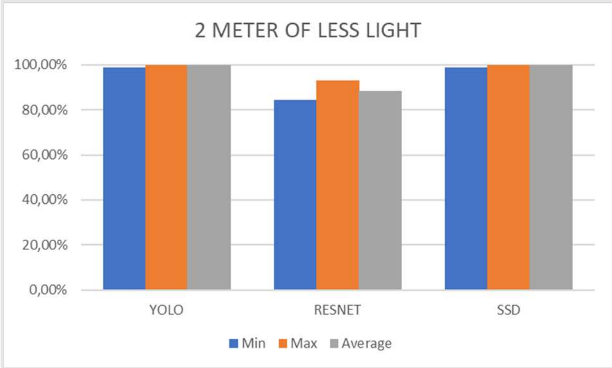

Fig 11. Results graph for 2 meter and less light

From the data results above, it can be seen that the best accuracy rate for a distance of 2 meters with less light is Yolo with an accuracy rate of 99.88%.

E. 3 Meter distance and normal light

The fifth experiment was conducted indoors with an object distance of 3 meters and normal lighting so as to get the accuracy level of each model with the same object, the results obtained are as follows:

TABLE 6. Accuracy results for a distance of 3 meter and normal light

	YOLO	RESNET	SSD
Min	50.17%	62.95%	61.59%
Max	99.96%	82.31%	99.79%
Average	90.26%	73.37%	94.13%

The figure below shows the findings of the comparison of the three models' degrees of accuracy:

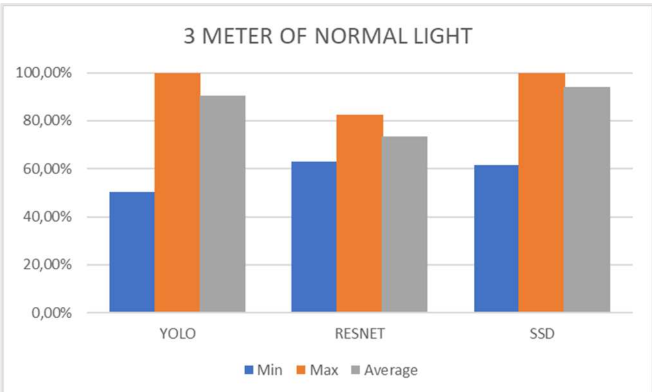

Fig 12. Results graph for 3 meter and normal light

From the data results above, it can be seen that the best accuracy rate for a distance of 3 meters with normal light is SSD with an accuracy rate of 94.13%.

F. *2 Meter distance and less light*

The sixth experiment was conducted indoors with an object distance of 3 meters and less lighting so as to get the accuracy level of each model with the same object, the results obtained are as follows:

TABLE 7. Accuracy results for a distance of 3 meter and less light

	YOLO	RESNET	SSD
Min	95.81%	68.60%	79.04%
Max	100.00%	81.39%	99.23%
Average	99.07%	74.94%	96.86%

The figure below shows the findings of the comparison of the three models' degrees of accuracy:

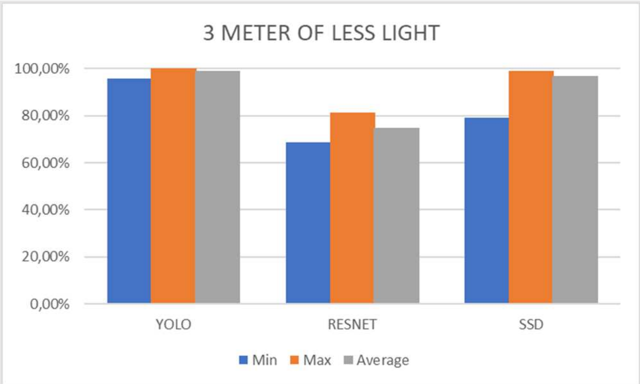

Fig 13. Results graph for 3 meter and less light

From the data results above, it can be seen that the best accuracy rate for a distance of 3 meters with less light is Yolo with an accuracy rate of 99.07%.

G. *Effect of Distance and Light*

To be able to determine the level of accuracy based on the influence of distance and light intensity, it can be seen in the following comparison table:

TABLE 8. Effect of distance and light intensity

	YOLO	RESNET	SSD
1M of Normal Light	99.67%	96.01%	98.82%
1M of Less Light	100.00%	96.14%	99.85%
2M of Normal Light	99.65%	85.02%	99.70%
2M of Less Light	99.88%	88.55%	99.78%
3M of Normal Light	90.26%	73.37%	94.13%
3M of Less Light	99.07%	74.94%	96.86%
Average	**98.10%**	**85.67%**	**98.19%**

The effect of distance and light on the comparison of the accuracy of the three models can be seen in the Fig 14 below:

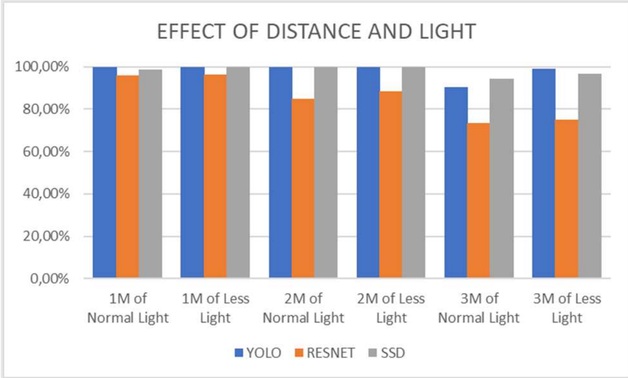

Fig 14. Accuracy based on distance and light intensity

Based on the results of the accuracy level in table 8 about the comparison between distance and overall light intensity, it can be concluded that the best accuracy of the average value for the three models with a distance of 1 - 3 meters is SSD with an average accuracy level of 98.19%. However, the Yolo model from 6 combinations of trials conducted there are 4 trials whose average accuracy level is higher than SSD. The overall average accuracy of SSD is higher because at a distance of 3 meters of normal light the Yolo model's accuracy level is unstable, so the average value of the distance decreases. Yolo has a high object detection speed, this is also in accordance with the test results that have been carried out by previous researchers [22] [23]. The accuracy level of Resnet is lower compared to YOLO and SSD because in Resnet the dimensions are changed in each block with the help of filters. Thus, jump connections face the problem of simply adding the input of the previous layer to the output of the next layer. To overcome this problem, the residual can be multiplied by a linear projection to align the dimensions.

For the results of the accuracy level based on normal and low light intensity, it can be concluded that images with 270 Lux light intensity (low light) have a greater accuracy level compared to normal light.

Based on the distance of taking objects with the camera, the results obtained are very influential, the farther the distance of taking objects, the accuracy level decreases, therefore to produce a good level of accuracy the distance needs to be considered, namely at a distance of 1 or 2 meters.

V. Conclusion

This article proposes a best model for object detection, especially person objects. Objects are identified with the help of bounding boxes. Experiments were conducted using three models such as Yolo, Resnet and SSD. The SSD model can perform object detection with the best accuracy in terms of distance and light intensity with an average value of 98.19%.

Image results generated from a room with a light intensity of 270 Lux produce a higher level of accuracy compared to normal light. In addition, the distance between the object and the camera greatly affects the results of the accuracy level.

For future research, it is expected that experiments can be carried out by testing the three models with facial expression datasets.

VI. References

[1] D. Zhou *et al.*, "IoU Loss for 2D/3D Object Detection," Aug. 2019, [Online]. Available: http://arxiv.org/abs/1908.03851

[2] H. Zhang, F. Shao, X. He, Z. Zhang, Y. Cai, and S. Bi, "Research on Object Detection and Recognition Method for UAV Aerial Images Based on Improved YOLOv5," *Drones*, vol. 7, no. 6, p. 402, Jun. 2023, doi: 10.3390/drones7060402.

[3] S. Singh, U. Ahuja, M. Kumar, K. Kumar, and M. Sachdeva, "Face mask detection using YOLOv3 and faster R-CNN models: COVID-19 environment," *Multimed Tools Appl*, vol. 80, no. 13, pp. 19753–19768, May 2021, doi: 10.1007/s11042-021-10711-8.

[4] M. Zahrawi and K. Shaalan, "Improving video surveillance systems in banks using deep learning techniques," *Sci Rep*, vol. 13, no. 1, p. 7911, Dec. 2023, doi: 10.1038/s41598-023-35190-9.

[5] Y. Li, J. Wang, J. Huang, and Y. Li, "Research on Deep Learning Automatic Vehicle Recognition Algorithm Based on RES-YOLO Model," *Sensors*, vol. 22, no. 10, May 2022, doi: 10.3390/s22103783.

[6] M. A. Moreno-Armendáriz, H. Calvo, C. A. Duchanoy, A. Lara-Cázares, E. Ramos-Diaz, and V. L. Morales-Flores, "Deep-learning-based adaptive advertising with augmented reality," *Sensors*, vol. 22, no. 1, Jan. 2022, doi: 10.3390/s22010063.

[7] M. Li, Z. Zhang, L. Lei, X. Wang, and X. Guo, "Agricultural greenhouses detection in high-resolution satellite images based on convolutional neural networks: Comparison of faster R-CNN, YOLO v3 and SSD," *Sensors (Switzerland)*, vol. 20, no. 17, pp. 1–14, Sep. 2020, doi: 10.3390/s20174938.

[8] T. Liu, B. Pang, S. Ai, and X. Sun, "Study on visual detection algorithm of sea surface targets based on improved yolov3," *Sensors (Switzerland)*, vol. 20, no. 24, pp. 1–14, Dec. 2020, doi: 10.3390/s20247263.

[9] A. Tellaeche Iglesias, I. Fidalgo Astorquia, J. I. Vázquez Gómez, and S. Saikia, "Gesture-Based Human Machine Interaction Using RCNNs in Limited Computation Power Devices," *Sensors (Basel)*, vol. 21, no. 24, Dec. 2021, doi: 10.3390/s21248202.

[10] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," Jun. 2015, [Online]. Available: http://arxiv.org/abs/1506.02640

[11] C. L. Lin and K. C. Wu, "Development of revised ResNet-50 for diabetic retinopathy detection," *BMC Bioinformatics*, vol. 24, no. 1, p. 157, Dec. 2023, doi: 10.1186/s12859-023-05293-1.

[12] I. Haider, H. J. Yang, G. S. Lee, and S. H. Kim, "Robust Human Face Emotion Classification Using Triplet-Loss-Based Deep CNN Features and SVM," *Sensors*, vol. 23, no. 10, May 2023, doi: 10.3390/s23104770.

[13] W. Liu *et al.*, "SSD: Single Shot MultiBox Detector," Dec. 2015, doi: 10.1007/978-3-319-46448-0_2.

[14] N. Sukhla, "A Review on Image Based Target Distance & Height Estimation Technique Using Laser Pointer and Single Video Camera for Robot Vision," 2015. [Online]. Available: www.researchpublish.com

[15] R. Hendrawan and W. Istiono, "Analyzing the Distance and Intensity of Light in Learning Augmented Reality Marker Based Tracking Application," *Journal of Advances in Mathematics and Computer Science*, pp. 58–70, Feb. 2023, doi: 10.9734/jamcs/2023/v38i21746.

[16] M. Ş. Gündüz and G. Işık, "A new YOLO-based method for social distancing from real-time videos," *Neural Comput Appl*, 2023, doi: 10.1007/s00521-023-08556-3.

[17] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," Jun. 2015, [Online]. Available: http://arxiv.org/abs/1506.02640

[18] T. Hong Chun *et al.*, "Efficacy of the Image Augmentation Method using CNN Transfer Learning in Identification of Timber Defect," 2022. [Online]. Available: www.ijacsa.thesai.org

[19] F. Ramzan *et al.*, "A Deep Learning Approach for Automated Diagnosis and Multi-Class Classification of Alzheimer's Disease Stages Using Resting-State fMRI and Residual Neural Networks," *J Med Syst*, vol. 44, no. 2, Feb. 2020, doi: 10.1007/s10916-019-1475-2.

[20] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement," Apr. 2018, [Online]. Available: http://arxiv.org/abs/1804.02767

[21] Badan Standardisasi Nasional, "Standar Nasional Indonesia Konservasi energi pada sistem pencahayaan," 2020, [Online]. Available: www.bsn.go.id

[22] L. Tan, T. Huangfu, L. Wu, and W. Chen, "Comparison of RetinaNet, SSD, and YOLO v3 for real-time pill identification," *BMC Med Inform Decis Mak*, vol. 21, no. 1, Dec. 2021, doi: 10.1186/s12911-021-01691-8.

[23] M. Ł. Kowalski *et al.*, "Detection of inflatable boats and people in thermal infrared with deep learning methods," *Sensors*, vol. 21, no. 16, Aug. 2021, doi: 10.3390/s21165330.