Real-time Social Distance Detection using YOLO-v5 with Bird-eye View Perspective to Suppress the Spread of COVID-19

1st Imam Husni Al Amin
Faculty of Information Technology and Industry
Universitas Stikubank
Semarang, Indonesia
imam@edu.unisbank.ac.id

2nd Falah Hikamudin Arby
Faculty of Information Technology and Industry
Universitas Stikubank
Semarang, Indonesia
falahhikamudin.arby@gmail.com

3rd Edy Winarno
Faculty of Information Technology and Industry
Universitas Stikubank
Semarang, Indonesia
edywin@edu.unisbank.ac.id

4th Budi Hartono
Faculty of Information Technology and Industry
Universitas Stikubank
Semarang, Indonesia
budihartono@edu.unisbank.ac.id

5th Wiwien Hadikurniawati
Faculty of Information Technology and Industry
Universitas Stikubank
Semarang, Indonesia
wiwien@edu.unisbank.ac.id

2022 2nd International Conference on Information Technology and Education (ICIT&E) | 978-1-6654-9433-5/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICITE54466.2022.9759552

***Abstract*—The COVID-19 virus outbreak has continued to spread since the end of 2019 worldwide. All people also implement health protocols not to contract this disease. One of the health protocols that must be implemented is to limit interactions between humans to a length of 1-2 meters or what is usually done with social distancing. Social distance detection system to ensure that people do not violate social distancing could be a solution to this problem. Using the YOLO-v5 method, which is the latest version of the YOLO (You Only Look Once) method with a detection speed of up to 140 Frames Per Second (FPS) and 90 percent smaller than the previous version, this system detects people who violate social distancing and then gives a voice warning to keep their distance to avoid spreading the COVID-19 virus. The human detection rate in the detection system reaches 93,5%, and the accuracy for social distance detection reaches 95%. Based on the research that has been done, it can be said that this system can work well for detecting social distance, but the detection will start detecting the distance between the camera and the object exceeding 10 meters.**

Keywords—artificial intelligence, computer vision, deep learning, social distancing, YOLO

I. Introduction

The increasing number of active cases of the Coronavirus 2019 (COVID-19) outbreak which continues to increase, has caused the death toll to continue to fall due to this virus. As time goes on, medical care units are overwhelmed by the increasing number of active cases of the outbreak. The World Health Organization (WHO) reports that on December 14, 2021, cumulative active cases reached 268,934,575 people, with 5,297,850 deaths worldwide [1]. The virus that spreads worldwide was identified as first appearing in Wuhan, China, at the end of 2019 [2]. With many active cases worldwide, the World Health Organization (WHO) has constantly been warning the whole world to be aware of the COVID-19 virus.

The COVID-19 virus has a high level of infection, which allows it to spread to others. The most common way for this virus to spread is through liquid droplets that assault the respiratory tract [3]. As a result, governments around the world, including in Indonesia, are asking citizens to follow health procedures such as wearing masks when traveling, washing hands often, and enforcing physical and social separation.

Social distance is one of the steps done by the government to reduce the risk of the COVID-19 virus spreading. The desirable social constraints when dealing with other people are not shaking hands and maintaining a distance of at least 1-2 meters [4]. The goal is to prevent being contaminated by liquid droplets from someone who is highly susceptible to infection. COVID-19. However, many people do not follow the guidelines of this social distance health plan in practice. When conversing with others, many people fail to notice their distance. This has been a source of concern for both the government and other communities, as it has the potential to cause harm.

Deep learning has become a hot topic in recent years, with applications ranging from object identification to facial recognition to various sorts of detection. Several well-known detection approaches, including Fast-RCNN, Faster-RCNN, SSD (Single Shoot Multibox Detector), and You Only Look Once (YOLO), have evolved into precise but lightweight detection systems that may be employed in a variety of applications [5]. According to research done in the form of person detection utilizing the YOLO method employed in the Smart Car system or Advanced Driver Assistance System, the accuracy of the human detection system may be relied on in Real Time employing 7 layers of Convolutional Neural Networks (ADAS) [6]. The detection of the employment of masks using the adapted YOLO-v4 approach is the subject of the following study. By improving the CSPDarkNet53 model, picture scaling algorithm, PANet structure, and data set based on mask wearing standards, it may address the problem of low accuracy and real-time performance in detecting people who wear masks or don't wear masks [7].

As a result, the goal of this research is to create a social distance detection system that can detect persons who violate the law while they are close to or interact with other people less than 1-2 meters away utilizing the YOLO method. This study is unique in that it incorporates a detection idea in the form of bird-view perspective measurements, ensuring that the detection results are more accurate when seen from the perspective of the image taken by the camera. The detection system will next be created using the YOLO-v5 model in conjunction with a real-time system.

Using the fifth version or the latest method of YOLO, the system will detect more quickly and accurately at longer distances than the previous version of YOLO. Then the system will also calculate the distance of each person detected in real-time using the Euclidean distance algorithm. So if someone violates social distancing, the system will give a warning in the form of a warning sound so that person immediately stays away from other people. The result of this study can encourage the public to implement health protocols such as social distancing, especially in public places and crowds, or any activity that has the potential to spread the COVID-19 virus, in order to prevent the virus from spreading both in Indonesia and around the world.

II. Method

A. *You Only Look Once (YOLO)*

YOLO is a deep learning method for detecting things that combines the detected object's components into a single neural network in the full image. The YOLO method's technology, namely YOLO, separates the input image into S × S sections or grids . When the center of an object falls within a grid cell, the grid cell is the one that detects the object. The bounding box and confidence are predicted by each grid cell. The confidence value indicates how sure a model is in the box containing the object, as well as how accurate the box is predicted. The equation is used to calculate the value of confidence:

$$Conf = \Pr(Object)\ x\ 10U_{pred}^{truth} \qquad (1)$$

The chance of an object arriving in a region is represented by Pr(object). In the meantime, IOU (Intersection Over Union) is the proportion of the ground truth and prediction box areas that overlap. As a result, the higher the number of IOUs (Intersection Over Union), the more precise the detection. If the confidence value is 0, it means there are no items in that cell. Every bounding box has five predictions: x, y, w, h, and confidence. The coordinates of the midpoint in relation to the grid cell boundaries are represented by the x and y values. The values of w and h, on the other hand, describe the width and height of the image as a whole. The Pr(Classi|Object) class probability value is also predicted by each grid. For each bounding box, YOLO calculates the class confidence value by multiplying the class probability value by the bounding box's confidence value. This number demonstrates how well the bounding box anticipates an object by indicating the probability class that appears in the bounding box [8].

B. *YOLO*-v5

As time goes by, technology is growing. Likewise with YOLO which has experienced many developments to be faster, stronger and better than the initial version. For now, the YOLO series includes YOLO-v1, YOLO-v2, YOLO-v3, YOLO-v4 and YOLO-v5. YOLO-v5 is the most recent version of the YOLO project. YOLO-v4 was used to create this. The YOLO-processing v5's speed has been greatly boosted, with the fastest speed hitting 140 frames per second (FPS). Because YOLO-v5 is so compact, even 90% smaller than YOLO-v4, it can be installed on an embedded device. A higher level of accuracy and the capacity to distinguish small things are other advantages. With the development of YOLO, the YOLO-v5 detection system managed to cover the shortcomings with its initial version, namely YOLO-v1. YOLO-v1 was developed based on the R-CNN method. Region Based CNN method or commonly called R-CNN.

The system of YOLO-v1 is that the input image is processed only once, hence it is called You Only Look Once. Then the different features are extracted through multiple convolution layers, and the convolution kernel parameters are shared each time. YOLO- v1 has a very fast image detection speed. It outperforms the previous detection model, R-CNN, in terms of speed and can meet real-time needs. The pace can approach 45 frames per second when the entire image is used as environmental data. However, the disadvantage of YOLO-v1 is that its position detection accuracy is low and small objects cannot be detected [9].

C. *Network* Structure Model

YOLO-v5 has a structure that is identical to previous YOLO versions. There are four sections to this structure: Backbone, Neck, and Prediction are all words that come to mind when thinking about input. Figure 1 adopted by previous research will explain the Architecture of the YOLO-v5 model [10].

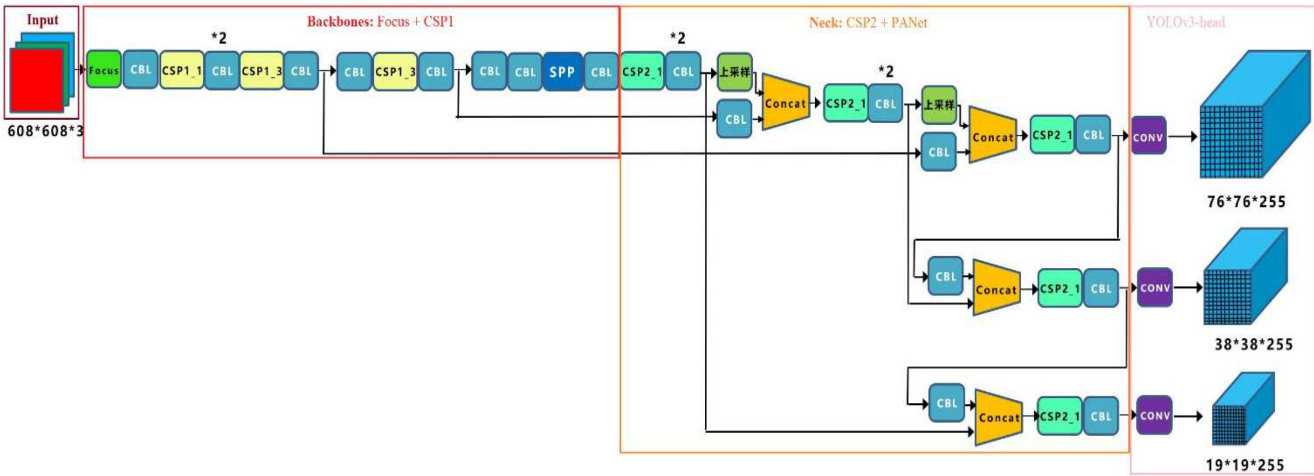

Fig. 1. Architecture of YOLO-v5 model

The input section allows for data augmentation, adaptive anchor frame calculation, and adaptive image scaling. The major objective of the feature extraction section is to use a focus structure to complete the cutting and convolution operations, as well as a CSP structure to improve the feature network's learning capability. The number of convolution kernels in the Focus and CBL of different networks varies, as does the number of residual modules in the CSP, the model's performance can be altered by varying the network's width and depth. The Neck part employs both the FPN and PAN structures, as well as information gleaned from the Backbone section, to enhance the network's overall capabilities. To offer predictive results, the output layer is separated into three convolution layer channels, which are calculated using the loss function, and the results are subjected to maximum value suppression processing [11].

D. Dataset

In the YOLO-built-in v5 dataset, the Common Objects in Context (COCO) dataset is pre-trained. COCO is a dataset that includes object recognition, segmentation, and labeling. In this collection, almost 200,000 pictures have been categorized into 80 different classifications, including the human class [12]. As a result, YOLO-v5 may be used to identify social separation, with the main system detecting humans first. Because the COCO dataset has already been trained for person detection, and the COCO dataset is the default dataset for YOLO-v5, there is no need to build a custom dataset first.

E. Research Stage

The YOLO-v5 approach has several stages for detecting persons who violate health norms in the form of social distancing, which occurs when someone interacts with others at a distance of less than 1-2 meters. Figure 2 is a flow of how the system will work. In Figure 2, it is explained that there are four core processes for this system. The first thing this system will do is calibrate the camera using a bird-eye view perspective.

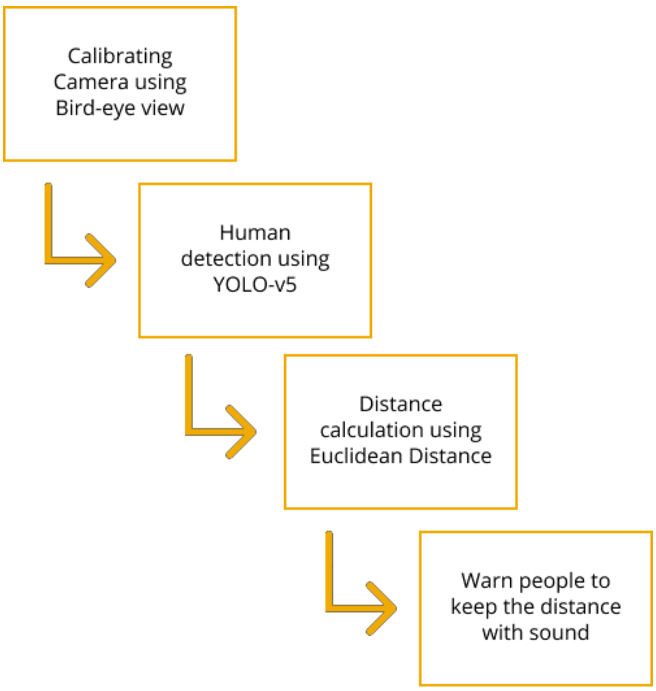

Fig. 2. System core process

The system will capture the frame first, and then we need to set 4 coordinate points that will be used as a benchmark for changing the perspective to a bird-eye view. With these coordinate points, the system will perform image processing to change the image's perspective to a bird-eye view. The frame will be stretched into a full rectangular frame to be further marked with a regular circle. After that, the size and range of the image frame will be returned as all and the circle mark. So that the circle will follow the perspective according to the coordinates that have been set previously, as described in more detail in Figure 3.

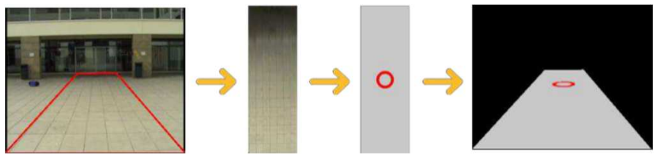

Fig. 3. Bird-eye view perspective image processing

After the image is processed into a bird-eye view perspective, it will create a variable for a circular sign that will be used as a human detection sign. The mark will adjust to every object that has been detected as a human. Then the image will perform image processing to return the display to its original state by bringing the mark that has followed the bird-eye view perspective. The next process is human detection using YOLO-v5. The camera will be turned on continuously to capture images in real-time. YOLO-v5 will process every frame captured by the camera to detect humans. Humans will be indicated with a bounding box that has been turned into a circle by altering the perspective of the bird's eye view in the preceding procedure.

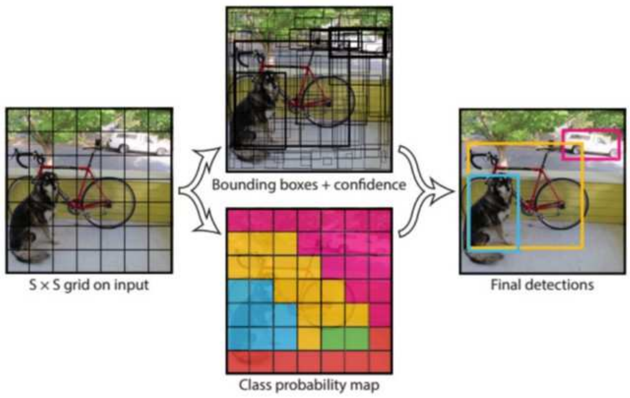

Fig. 4. YOLO object detection system model [13]

YOLO will divide the system into S x S grids on the input. Then each grid cell will be made up of bounding boxes, confidences for each box, and probability classes (see Figure 4). So that in the end it will produce object detection predictions for each class. After every human is detected in the system, the next step is to calculate the distance of each human using the Euclidean Distance formula. To examine the link between angles and distances, this formula is used to compute the distance between two points in Euclidean space. The Euclidean distance between two data points x and y in d-dimensional space is the most often used distance computation formula for numerical data [14].

The Euclidean distance (d) can be calculated using the equation for two-dimensional calculations at the coordinates (x1,y1) and (x2,y2).

$$d\ (P1P2) = \sqrt{(Y_1 - Y_1)^2 + (Y_2 - Y_2)^2} \qquad (2)$$

Then the last process is the sound output that will be issued in the form of a warning if there are two or more humans adjacent to a distance of less than 1-2 meters so that the circle marks on the system clash. Collision signs will turn red as a sign that the system has caught a social distancing violation. The sound emission is designed to notify social distancing offenders to retain their distance in order to limit the risk of getting the COVID-19 virus.

III. Result and Discussion

The outcomes of the social distancing detection system that has been built will be tested and analyzed in this study. The experiment was carried out with an Android smartphone camera connected to a laptop via the DroidCam application in order to capture photos more freely in real-time. Previous research has been conducted using artificial intelligence-based smart camera systems [15]. Furthermore, because the dataset utilized is the default dataset from YOLO-v5, namely COCO Dataset, no new dataset training is required because COCO Dataset already has data for person detection that has been trained on over 200,000 photos [11], [16]. Using YOLO-v5, this test can be used to determine the amount of success of this system in detecting social distancing violations. The formula is used to calculate the success rate of testing on this system:

$$k = \frac{n}{m}\ x\ 100\% \qquad (3)$$

Where:

k = percentage of test success
n = number of successful detections
m = number of observed data

The percentages of two forms of detection are calculated in the test results: the percentage of human detection using Yolo-v5 and the percentage of social distancing detection, whether it has been detected or not.

Fig. 5. Test image 1

The first experiment was conducted with two people interacting, as shown in Figure 5. It appears that the two people interacted with a distance of less than 1-2 meters, so that they were violating social distancing. With a distance of 4 meters from the object to the camera, the system achieves an accuracy level of 100% for human detection using Yolo-v5 because it can detect the two object (people) and give a circle mark on each object [17]–[19]. The system also managed to reach a 100% level for social distancing detection accuracy. It is because the system managed to detect that the two objects violated social distancing by changing the color of the circle sign from green to red and issuing a warning sound to keep the distance.

Fig. 6. Test image 2

The second study, conducted on a highway with two human objects shown in Figure 6. Two people appear to be sweeping the road, yet they keep their distance from one another. With a camera-to-object distance of 7 meters, the system achieved an accuracy rate of 100% for human detection using YOLO-v5. Because it looks like the two people are keeping a distance from each other, the accuracy rate for social distancing detection is 100% because the system does not identify the two persons breaking social distancing by placing a green mark on the circle for each person.

Fig. 7. Test image 3

Then for the third study was carried out with many people, according to Figure 7. Here the distance from the camera to the object has various distances because there are many people. The furthest distance reaches 8 meters. The system succeeded in detecting six humans caught by the camera, so the accuracy for the human detection system using YOLO-v5 in the third study reached 100%. The level of accuracy for the social distancing detection system also reaches 100%. The detection result reaches 100% because the system successfully detects that four people do not keep their distance by providing a red circle indicator.

The fourth study was performed with multiple persons on the highway, and the camera distance was extremely large, as shown in Figure 8. The test was conducted with the camera a few meters away from the object, which reached a distance of 10 meters with some persons walking and others not keeping their distance.

Fig. 8. Test image 4

The system detected six people, but seven people were caught on camera. That is because other objects somewhat covered that one person, so the camera did not catch it clearly, So, in this third test, the success percentage for human detection using YOLO-v5 was 85.7%. For social distancing detection, it reaches 100% because the system has succeeded in detecting people who do not keep the distance caught by the camera.

Fig. 9. Test image 5

The camera employed an angle from above, such as CCTV, the fifth investigation was conducted at a large distance (as seen in Figure 9). The camera recorded many pedestrians, some of whom did not keep their distance. The distance for the camera to the farthest object is 15 meters. The accuracy rate for the human detection system using YOLO-v5 reaches 86.3% because some are not detected. Because the system fails to detect accurately, it appears that someone has kept their distance, but instead, it is identified that they are breaching social distancing, the accuracy of social distancing detection approaches 75%. Table 1 summarizes the comparison of all the results of the analysis on the tests that have been conducted.

Table 1 shows that with a camera distance of 4 to 15 meters, the system can detect human objects and detect social distancing violations with an average accuracy of 93.5% for detecting persons and 95% for detecting social distancing violations. The accuracy of social distancing is shown to be dependent on human detection accuracy in this experiment. The detection of social separation can only be quantified if a human is observed. While human detection accuracy with YOLO-v5 is quite high while the object is visible, it is ineffective when the object is too far away [20][22]. As a result, it appears little to the camera, and it is unable to recognize when a human is stacked with another object, preventing the camera from seeing it clearly.

The social distancing detection system that has been developed has been proven to work well for detecting social distancing. Using the dataset from YOLO-v5, particularly the COCO dataset, the system can detect humans well by default. The level of accuracy obtained for human detection accuracy using YOLO-v5 reaches 93.5%. The YOLO-v5 person detection technology has one flaw: if the distance between the camera and the object exceeds 10 meters, the accuracy level gradually decreases. Euclidean distance is used to calculate the distance from two objects so that the system can detect social distancing violations [23]. The accuracy of detecting social distancing violations has reached 95%. The results of this study may aid community and government efforts to combat the COVID-19 virus's proliferation.

TABLE I. SYSTEM ANALYSIS PERCENTAGE RESULTS

Test Image	Farthest Object Distance	Human Detection	Social Distancing Detection
1	4 m	100%	100%
2	7 m	100%	100%
3	8 m	100%	100%
4	10 m	85.7%	100%
5	15 m	81.8%	75%
Average		93.5%	95%

IV. CONCLUSION

After conducting research on the social distance detection system, it has been determined that it is capable of detecting social distancing infractions. The system does not need to be trained for person detection because the dataset from YOLO-v5, notably the Common Objects in Context (COCO) dataset, can detect humans well by default. The level of accuracy obtained for human detection accuracy using YOLO-v5 reaches 93.5%. This occurs because the YOLO-v5 person detection system has a flaw: if the distance between the camera and the object surpasses 10 meters, the accuracy level gradually decreases. Otherwise, the object looks small in the camera, and the system cannot detect if another object is blocking the human. To detect social distancing breaches, the system uses Euclidean distance to determine the distance between two human objects. The accuracy of identifying violations of social separation has achieved a satisfactory level of 95%. This research can help the government suppress the spread of the COVID-19 by detecting people closer than 1−2 meters. The system will inform that a social distancing violation has occurred so that people who violate it will be aware and not too close to other people, as a result, the COVID-19 virus will not spread.

REFERENCES

[1] World Health Organization (WHO), "COVID-19 weekly epidemiological update," World Heal. Organ., no. 58, pp. 1–23, 2021,

[Online]. Available: https://www.who.int/publications/m/item/covid-19-weekly- epidemiological-update

[2] A. R. Fahindra and I. H. Al Amin, "Sistem Pakar Deteksi Awal Covid-19 Menggunakan Metode Certainty Factor," J. Tekno Kompak, vol. 15, no. 1, pp. 92–103, 2019, doi: 10.33365/jtk.v15i1.914.

[3] Q. Han, Q. Lin, Z. Ni, and L. You, "Uncertainties about the transmission routes of 2019 novel coronavirus," Influenza Other Respi. Viruses, vol. 14, no. 4, pp. 470–471, 2020, doi: 10.1111/irv.12735.

[4] A. Kresna and J. Ahyar, "Efektivitas Physical Distancing dan Social Distancing Terhadap Kesehatan Dalam Pendekatan Linguistik," J. Syntax Transform., vol. 1, no. 4, pp. 14–19, Jun.2020, doi: 10.46799/%j.vol1.iss4.42.

[5] L. I. N. Wang, "Tinier-YOLO : A Real-Time Object Detection Method for Constrained Environments," IEEE Access, vol. 8, pp. 1935–1944, 2020, doi: 10.1109/ACCESS.2019.2961959.

[6] M. H. Putra, Z. M. Yussof, S. I. Salim, and K. C. Lim, "Convolutional neural network for person detection using YOLO framework," J. Telecommun. Electron. Comput. Eng., vol. 9, no. 2–13, pp. 1–5, 2017.

[7] J. Yu and W. Zhang, "Face Mask Wearing Detection Algorithm Based on Improved YOLO-v4," Sensors 2021, Vol. 21, Page 3263, vol. 21, no. 9, p. 3263, May 2021, doi: 10.3390/S21093263.

[8] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., vol. 2016- Decem, pp. 779–788, 2016, doi: 10.1109/CVPR.2016.91.

[9] Y. Liu, B. Lu, J. Peng, and Z. Zhang, "Research on the Use of YOLOv5 Object Detection Algorithm in Mask Wearing Recognition," vol. 6, no. 11, pp. 276–284, 2020, doi: 10.6911/WSRJ.202011.

[10] M. Chablani, "YOLO — You only look once, real time object detection explained," Towards Data Science, 2017. https://towardsdatascience.com/yolo-you-only-look-once-real-time-object-detection-explained-492dc9230006 (accessed Nov. 11, 2021)

[11] J. Zhou, P. Jiang, A. Zou, X. Chen, and W. Hu, "Ship Target Detection Algorithm Based on Improved YOLOv5," 2021.

[12] M. Kasper-eulaers, N. Hahn, S. Berger, T. Sebulonsen, and Ø. Myrland, "Short Communication : Detecting Heavy Goods Vehicles in Rest Areas in Winter Conditions Using YOLOv5," 2021.

[13] V. A. Adibhatla, H. C. Chih, C. C. Hsu, J. Cheng, M. F. Abbod, and J. S. Shieh, "Applying deep learning to defect detection in printed circuit boards via a newest model of you-only-look-once," Math. Biosci. Eng., vol. 18, no. 4, pp. 4411–4428, 2021, doi: 10.3934/mbe.2021223

[14] N. N. Nurliza, "Penerapan Euclidean Distance Pada Pengenalan Pola Citra Sidik Jari," pp. 1–67, 2018.

[15] O. Karaman, A. Alhudhaif, and K. Polat, "Development of smart camera systems based on artificial intelligence network for social distance detection to fight against COVID-19," Appl. Soft Comput., vol. 110, p. 107610, 2021, doi: https://doi.org/10.1016/j.asoc.2021.10761

[16] H. Caesar, J. Uijlings, and V. Ferrari, "Coco-stuff: Thing and stuff classes in context," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 1209–1218.

[17] Z. Wang, Y. Wu, L. Yang, A. Thirunavukarasu, C. Evison, and Y. Zhao, "Fast personal protective equipment detection for real construction sites using deep learning approaches," Sensors, vol. 21, no. 10, p. 3478, 2021.

[18] G. Yang et al., "Garbage Classification System with YOLOV5 Based on Image Recognition," in 2021 IEEE 6th International Conference on Signal and Image Processing (ICSIP), 2021, pp. 11–18.

[19] T. Jintasuttisak, E. Edirisinghe, and A. Elbattay, "Deep neural network-based date palm tree detection in drone imagery," Comput. Electron. Agric., vol. 192, p. 106560, 2022.

[20] F. Zhou, H. Zhao, and Z. Nie, "Safety helmet detection based on YOLOv5," in 2021 IEEE International Conference on Power Electronics, Computer Applications (ICPECA), 2021, pp. 6–11.

[21] X. Zhu, S. Lyu, X. Wang, and Q. Zhao, "TPH-YOLOv5: Improved YOLOv5 Based on Transformer Prediction Head for Object Detection on Drone-captured Scenarios," in Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 2778–2788.

[22] B. Yan, P. Fan, X. Lei, Z. Liu, and F. Yang, "A real-time apple targets detection method for picking robot based on improved YOLOv5," Remote Sens., vol. 13, no. 9, p. 1619, 2021.I. Dokmanic, R. Parhizkar,

[23] J. Ranieri, and M. Vetterli, "Euclidean distance matrices: essential theory, algorithms, and applications," IEEE Signal Process. Mag., vol. 32, no. 6, pp. 12–30, 2015..