Literasi Nusantara
vol.5 no. 3, July 2025 – October 2025
https://journal.citradharma.org/index.php/literasinusantara/index

ISSN 2746 8208 (Online)
ISSN 2746 3575 (Print)

Implementation of Yolov8 to Detect Focus and Fatigue Levels of Employees

Kevin Hanif Wicaksana [1], I Gede Susrama Mas Diyasa [2], *Fetty Tri Anggraeny [3]

[1,2,3] Universitas Pembangunan Nasional Veteran, Jawa Timur, Surabaya, Indonesia
kevinhanif27@gmail.com , igsusrama.if@upnjatim.ac.id ,
fettyanggraeny.if@upnjatim.ac.id

DOI : *https://doi.org/10.56480/jln.v5i2.1572*

Received: April 18, 2025
Revised: May 24, 2025
Accepted: June 04, 2025

Abstract

The coronavirus disease (Covid-19) pandemic that occurred at the end of 2019 had a major impact worldwide. In addition to impacting the health sector, other sectors such as business and technology are also very influential factors. At the same time, trends in artificial intelligence, machine learning, and deep learning are also emerging, bringing tremendous developments worldwide. New methods and advanced architectures such as YOLO (You Only Look Once) are attracting much attention because they can accurately detect objects with very high probability. YOLO is considered the "fastest deep learning object detector" and is an object detection network architecture that focuses on accuracy and speed. YOLOv8 allows you to perform accurate facial recognition to detect fatigue on employee faces. This detection evaluates how well employees focus on their work. To help companies maintain safety and prevent accidents at work. Such as jobs that require a level of focus to avoid accidents such as construction workers or taxi drivers or customers or people related to the job. The best level of accuracy obtained with YOLOv8 in real-time, has a detection accuracy or large mAP (Mean Average Precision) value of up to 0.977 or 97.7%.

Keywords– *Yolo, Fatigue Detection, Artificial Intelligence, Machine Learning.*

1. Introduction

Technological developments have led to the digitalization of work and changed the way people produce and interact. Global technological developments were first observed since the First Industrial Revolution in England in the early 18th century to the early 19th century. This first revolution was marked by the discovery of the steam engine and the use of coal energy. The discovery of this first technology succeeded in changing the economic structure of society from an agricultural base to an industrial base. Further developments occurred in the late 19th century, with the discovery of new energy sources: electricity, gas, and oil, marking the beginning of the Second Industrial Revolution initiated by the United States (Khan, 2024).

The coronavirus disease (Covid-19) pandemic that occurred at the end of 2019 had a major impact worldwide. In addition to impacting the health sector, other sectors such as business and technology are also very influential factors. At the same time, trends in artificial intelligence, machine learning, and deep learning have also emerged, bringing extraordinary developments worldwide (Bansal et al., 2020; Bontridder & Poullet, 2021).

Digitalization in the Industrial Revolution 4.0 era does not only mean the use of computer technology and the internet to communicate, but the internet is also used as a database that reflects market changes in developing countries. Digitalization makes work more flexible and gives employees greater autonomy without being limited by time or location. Digitalization can be interpreted as the process of digitizing everything that can be digitized and changing the format of information into a digital format. In business, digitalization allows all stages of production to be carried out using big data to obtain market information (Sun et al., 2025).

Digital changes in large-scale industries have an impact on the automation of production equipment. For example, production machines replace traditional (manual) machines with digital machines or AI (Artificial Intelligence). On the other hand, automation of production equipment requires more skilled workers, and even existing employees may not have the skills to use all digital tools. The

Literasi Nusantara. vol.5 no. 3, July 2025 – October 2025
ISSN 2746 8208 (Online) and ISSN 2746 3575 (Print)

impact is a reduction in the number of employees. Machine replacement will be carried out gradually to avoid significant personnel reductions (Bruno, 2024).

YOLOv8 is a type of neural network model for object detection, and is part of the "You Only Look Once" (YOLO) series of models. This series aims to achieve real-time object detection, accurately identify and locate objects in images, and describe bounding boxes and object categories (Tang & Guo, 2024). Built on YOLOv7, YOLOv8 improves detection accuracy and performance by refining network architecture, adjusting model parameters, and optimizing training strategies. Typically, it uses a Convolutional Neural Network (CNN) architecture, utilizing Anchor boxes and feature pyramid networks to detect objects of various sizes and aspect ratios (Tang & Guo, 2024).

YOLOv8 combines advanced techniques and strategies to achieve higher detection accuracy and faster processing speed. Introducing new feature extraction methods, improved loss functions, more efficient model structures, and effective training techniques, YOLOv8 has made significant progress and advancement in the field of object detection (Tang & Guo, 2024).

Therefore, researchers are trying to create a trial method that can detect worker fatigue in order to maintain the level of safety and security at work. This will certainly affect the efficiency of companies in selecting suitable workers according to the level of focus. For construction workers and drivers, of course, this technology is needed to maintain the safety of workers and customers. With the help of facial expression detection using YOLOv8, companies can find out the level of focus and fatigue in their employees. If the employee's facial expression is often sleepy, often unfocused and easily tired, then the employee must be watched out for because it is feared that it can cause accidents at work if they are not focused on working. Maintaining diet and sleep patterns are factors that can increase employee focus. If employees can focus and work optimally, of course the company will at least reduce the number of accidents in workers (Onososen et al., 2025).

2. Method

Convolutional Neural Network (CNN) is a deep learning method that has evolved from the multilayer perceptron (MLP). It currently delivers the most significant advancements in image recognition due to its ability to mimic the human visual cortex in processing visual data. On the other hand, YOLO (You Only Look Once) is an object detection algorithm that differs from conventional approaches by utilizing a single neural network to analyze the entire image in one go (Hussain, 2024).

CNN was initially introduced as the NeoCognitron by Kunihiko Fukushima, a scientist at the NHK Broadcast Science Research Institute located in Kinuta, Setagaya Ward, Tokyo. Later on, the concept was enhanced by Yann LeCun, a researcher from AT&T Bell Laboratories in Holmdel, New Jersey, USA. LeCun developed a CNN model known as LeNet, which proved effective in his studies involving digit and handwriting recognition. In 2012, Alex Krizhevsky leveraged a CNN-based model to win the ImageNet Large-Scale Visual Recognition Challenge, demonstrating the power of deep learning methods—particularly CNNs. Compared to traditional machine learning techniques like Support Vector Machines (SVM), CNN has shown superior performance in image classification tasks (Prince et al., 2024).

YOLO is an integrated convolutional neural network capable of predicting multiple bounding boxes and their corresponding class probabilities at the same time. It is trained end-to-end on entire images, enabling rapid optimization of detection accuracy. This unified approach offers several benefits compared to conventional object detection techniques (Vijayakumar & Vairavasundaram, 2024). You Only Look Once (YOLO) is a deep learning algorithm that applies convolutional neural networks (CNNs) for object detection tasks. The method divides an input image into an s x s grid, where each grid cell is responsible for predicting bounding boxes and associated class probabilities. When an object is present within a grid cell, the algorithm predicts a bounding box around it. Each bounding box is assigned a confidence

Literasi Nusantara. vol.5 no. 3, July 2025 – October 2025
ISSN 2746 8208 (Online) and ISSN 2746 3575 (Print)

score, and the final prediction is determined based on the highest confidence values (Malahina et al., 2024).

Here are the various parameters that can be set or configured in YOLO:

a. Batch size

Batch size is a variable that determines how many images or training data are input during training. The smaller the batch size, the faster the training process. On the other hand, the larger the value of the packet size, the more storage capacity is required, thus extending the training process. This also affects the accuracy of the system. A larger value of the packet size used results in higher accuracy because the system learns more features (Malahina et al., 2024).

b. Segmentation

The department divides the deployment cost into smaller units called mini-batches. Dividing it into 8 subsections using a batch value of 64 gives a value of 8. This means that the training process is completed for 8 images per minibatch. This process is repeated eight times until the learning process for one section is complete. The system then processes the next batch with a cost of 64. The segmentation process aims to speed up the learning process while increasing GPU accuracy (Sahafi et al., 2024).

c. Channels

The channel value determines the depth of the image based on the data used during the training process. If the trained data uses RGB images, the channel variable must be 3. On the other hand, when using grayscale images, the channel uses a value of 1 (He et al., 2024).

d. Learning Rate

The learning rate is a determinant of the weights updated during backpropagation errors. The learning rate also determines the rate of repetition to achieve the minimum loss function. The faster the learning process, the higher the learning speed. However, if the learning rate is too high, irregular fluctuations in the loss function value can occur, so several attempts are needed to obtain the optimal learning rate value (Elgamily et al., 2025).

e. Maximum number of batches

The maximum number of batches is the number of iterations of the training data process. The higher the maximum package value, the more the system will learn the training data. The amount of training data should not exceed the maximum number of packages. The maximum package value must be adjusted to the number of object classes to be detected (Malahina et al., 2024).

YOLOv8 is a type of neural network model for object detection and belongs to the You Only Look Once (YOLO) family of models. The goal of this series is to achieve real-time object detection, accurately identify and localize objects in images, and describe bounding boxes and object categories (Widianto, 2023).

Built on YOLOv7, YOLOv8 improves detection accuracy and performance by refining the network architecture, optimizing model parameters, and optimizing training strategies. These models typically use a convolutional neural network (CNN) architecture and utilize anchor boxes and feature pyramid networks to detect objects of various sizes and aspect ratios (Widianto, 2023).

YOLOv8 combines advanced techniques and strategies to achieve higher detection accuracy and faster processing speed. By introducing new feature extraction methods, improved loss functions, more efficient model structures, and effective training techniques, YOLOv8 represents a significant step forward and advancement in the field of object detection. The YOLOv8 network architecture consists of three main parts: backbone, neck, and head (Widianto, 2023).

Pre-Processing Data

In conducting this research, the following are the methods used for the research experiment, namely:

Literasi Nusantara. vol.5 no. 3, July 2025 – October 2025
ISSN 2746 8208 (Online) and ISSN 2746 3575 (Print)

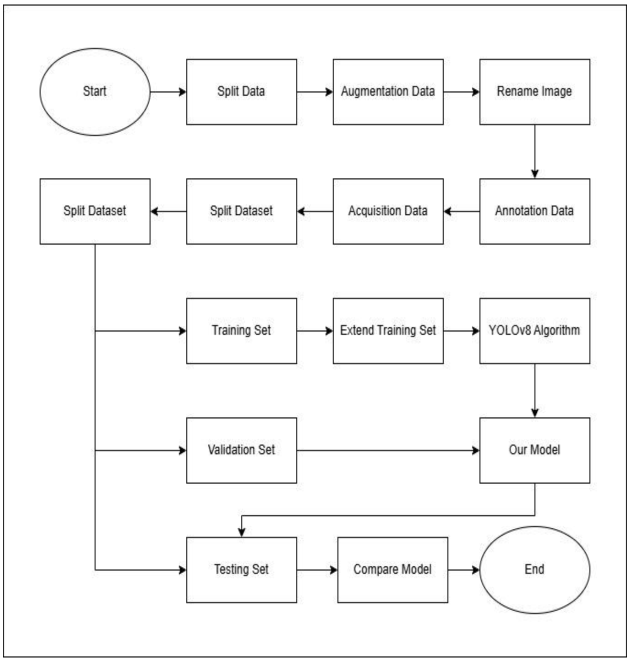

Figure 1. YOLOv8 System Workflow

In the data preprocessing section, the image data that has been collected is then processed, and the format changes according to the Yolo format. The format changes that need to be made are

a. Data division

Data division is the process of recording distributed data from "active" and "fatigue" data records. Separation of data records based on this class encourages researchers to expand, identify, and complete images.

b. Data augmentation

Data augmentation is a change in the shape of an image by mirroring the image, so that the resulting image is different from before. This is used to multiply image data in data records.

c. Data name changes

To make data grouping and labeling easier, image data renaming is required. Each class consists of 1,000 photos. So the entire image is in 2000 data records.

d. Data annotation

The image dataset that has been divided based on the previous class, then the image annotation process is carried out on the image according to the YOLO format using the Roboflow Web Application.

This stage is a new stage in annotating data. With Roboflow, dataset processing becomes simpler, faster, and easier to do. In the past, programs or applications used for data labeling or data annotation used labelimg. However, because the application needs to be installed, using roboflow is much simpler because it only requires a browser and internet connection.

e. Data acquisition

The annotated dataset is then processed by dividing the image images into training data, valid data, and test data. This data acquisition process is carried out using Roboflow. This aims to facilitate data acquisition randomly, quickly, and accurately. And it can also minimize errors made by researchers.

In this process, researchers divide the data into 3 parts, namely: 1. Train set or training data, 2. Valid set or valid data, and 3. Test set or test data. Each part has a different amount of data.

The size or comparison of the amount of data divided from the dataset, namely: 1. Train set or training data of 70% of the data or images from the dataset, 2. Valid set or valid data of 20% of the data or images from the dataset, and 3. Test set or test data of 10% of the data or images from the dataset.

Training Data

Training data is data used for the data training process or the process of training object recognition. Valid data is data used to validate the accuracy of the training data. Test data is data used to test the detection results from data training.

Data that has been pre-processed data, then will be trained to detect objects. The data training stage is as follows:

Literasi Nusantara. vol.5 no. 3, July 2025 – October 2025
ISSN 2746 8208 (Online) and ISSN 2746 3575 (Print)

a. YOLOv8 Installation

In starting the data training process, YOLOv8 is installed on Google Colaboratory. YOLOv8 used in this program is version 8.0.196. This version is the latest version of YOLOv8. And YOLOv8, created and developed by Ultralytics.

b. Pre-training process with the COCO model

This process is to prepare the YOLO model to be used. As an experiment, the default detection from YOLOv8 is used and using photo objects from Roboflow. Roboflow is a web platform that has functions related to dataset collections. Prediction and detection results will be stored in the 'runs/predict' folder.

c. Preparing Dataset

In preparing a customized dataset, there are several steps that must be taken. Roboflow is a web application that can be used to prepare custom datasets. By using Roboflow, this process is done by entering the API project dataset into Google Collaboratory.

d. Data Training

In this process, data training with YOLOv8 only requires 10-25 epochs to obtain the best detection accuracy results. Epoch is a repetition of a detection experiment to find the best detection configuration. The estimated time for training with 25 epochs is approximately 0.394 hours or around 23 minutes.

Training data and validation data come from an open-source image dataset from Kaggle that has been augmented in annotations. Training data and validation data must not have the same images. While the test data is the image that will be tested. The test data is obtained based on images obtained by researchers from the Internet and data created by themselves with fellow researchers.

During the training phase, the configured hyperparameter learning rate is relatively low, set at 0.001. As a result, the training process requires a longer time to complete. However, this small learning rate contributes to achieving higher accuracy. Therefore, in the event of an interruption or error during training, it is essential to retain the most recent weights so that the training can be resumed without starting over from scratch.

e. Saving Weights Files and Loading Trained Weights

This step plays a crucial role in the data training process. The trained data is saved in a file format known as a weights file (.weights). It is essential that this saved data can be reloaded, allowing the training to resume or the model to be used for further processing. In cases where an error occurs during training, the saved weights can be utilized to continue the training process without starting from the beginning.

f. Connecting to Google Drive

Connecting Google Colaboratory with Google Drive needs to be done so that during the detection trial process, photos of object detection results can be stored on Google Drive and are not lost when an error occurs.

Data Validation

Once the data training is successfully completed, the subsequent step involves evaluating the trained model. The following outlines the stages involved in the data evaluation process:

a. Calculating Image Detection

During this process, both the detected images and the total number of objects will be counted. In addition to the object count, several metrics such as the Average Precision (AP) value, the number of True Positives (TP), False Positives (FP), and the count of unique ground truths will also be calculated. The results from the true and false detections will be used to determine the accuracy level of the image detection.

b. Saving Output Results and Calculating Parameter Performance

Based on the detection results and image detection calculations during the training process, the test result data will be stored in a backup. This storage is important for tracking the progress in accuracy improvement. Performance parameter calculations are necessary to monitor whether accuracy is increasing or decreasing. This step provides

Literasi Nusantara. vol.5 no. 3, July 2025 – October 2025
ISSN 2746 8208 (Online) and ISSN 2746 3575 (Print)

preliminary insights into whether the accuracy achieved is suitable for use. The accuracy results are displayed in graphical form. However, an issue arises when displaying the accuracy graph, as the program is unable to show the performance from previous stages if there is an interruption during data training. As a result, the last weights file cannot be displayed.

3. Result and Discussion

After understanding the methodology employed for the subject focus level detection research, the next step is to implement the research method and present the data that supports the results of the conducted study. Based on the findings from the research, the following is a discussion of the results obtained from the study:

Data Training Results

The process of constructing the YOLO architecture for data training is relatively straightforward. The training process typically takes between 10 to 25 epochs, with an estimated duration of approximately 10 to 25 minutes.

Data Evaluation Results

Based on the results of the data evaluation obtained, the following is a graph of the results of the mAP value performance that has been successfully combined into 1 graph:

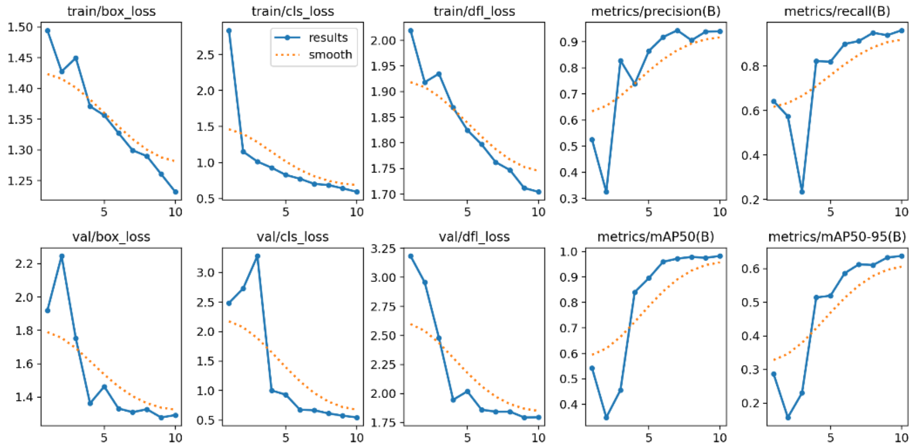

Figure 2. Data Training Result Graph

Based on the graph, it can be seen that the data training process went well and there were no obstacles. The precision value reached 0.94 or 94%. While the recall value reached 0.96 or 96%. The mAP50 value reached 0.98 or 98%. And the mAP50-95 value reached 0.63 or around 63%

Fatigue Detection Results

In the process of building the YOLO network architecture for detection through images, photos, and video streams, it takes quite a long time. The obstacles obtained are that sometimes errors occur during the process of taking photos and video streams. The final results obtained in fatigue detection are quite good and in accordance with expectations. Fatigue detection on the face can be detected well and there is an mAP value in the bounding box. So that fatigue detection or maintained, it can be seen how much mAP value is obtained. Here are some results of fatigue detection using photos from researchers:

Table 1. Focus and Fatigue Detection using YOLOv8

Detection Subject	**Detection Results**	Score
	Focus	0.65
	Not Focus	0.72
	Not Focus	0.78

Literasi Nusantara. vol.5 no. 3, July 2025 – October 2025
ISSN 2746 8208 (Online) and ISSN 2746 3575 (Print)

Based on the results of fatigue detection obtained from 3 selfie photos of researchers, the results obtained were 1 focused photo and 2 unfocused photos. Seen in the bounding box, it says "focus" and "not_focus" with the detection score on the right side of the label. For the first photo, a score of 0.65 or 65% was obtained. For the second photo, a score of 0.72 or 72% was obtained. And for the last photo, a score of 0.78 or 78% was obtained.

4. Conclusion

The results of the detection accuracy, measured by the mean Average Precision (mAP) value, often exhibit fluctuations during the training process, increasing or decreasing as the model iteratively learns from the data. However, these values tend to stabilize once the model reaches a sufficiently high number of epochs, typically in the range of 10 to 25 epochs. This stabilization indicates that the model has converged, meaning further training yields minimal improvements in performance. The consistent trend in mAP values suggests that the data training process is proceeding effectively without significant obstacles, such as overfitting or underfitting. Notably, the model achieves a high precision value of 0.94 (94%), demonstrating its ability to correctly identify positive instances with minimal false positives. Additionally, the recall value of 0.96 (96%) reflects the model's effectiveness in capturing the majority of true positive cases, minimizing false negatives. The mAP50, which evaluates detection accuracy at an Intersection over Union (IoU) threshold of 0.50, reaches an impressive 0.98 (98%), indicating near-perfect performance under this metric. Meanwhile, the mAP50-95, which averages precision across IoU thresholds ranging from 0.50 to 0.95 in increments of 0.05, achieves a value of 0.63 (63%), highlighting the model's robustness across varying levels of detection strictness. Furthermore, in a practical experiment involving three researcher-supplied photos, the model demonstrates exceptional real-world applicability, achieving 100% accuracy with mAP scores ranging between 0.68 and 0.78. These results underscore the model's reliability and adaptability in diverse detection scenarios, making it a promising tool for applications requiring high precision and recall.

References

Bansal, A., Padappayil, R. P., Garg, C., Singal, A., Gupta, M., & Klein, A. (2020). Utility of Artificial Intelligence Amidst the COVID 19 Pandemic: A Review. *Journal of Medical Systems*, *44*(9). https://doi.org/10.1007/s10916-020-01617-3

Bontridder, N., & Poullet, Y. (2021). The role of artificial intelligence in disinformation. *Data and Policy*, *3*(3), 1467–1497. https://doi.org/10.1017/dap.2021.20

Bruno, Z. (2024). The Impact of Artificial Intelligence on Business Operations. *Global Journal of Management and Business Research*, *9*(9), 1–8. https://doi.org/10.34257/gjmbrdvol24is1pg1

Elgamily, K. M., Mohamed, M. A., Abou-Taleb, A. M., & Ata, M. M. (2025). Enhanced object detection in remote sensing images by applying metaheuristic and hybrid metaheuristic optimizers to YOLOv7 and YOLOv8. In *Scientific Reports* (Vol. 15, Issue 1). https://doi.org/10.1038/s41598-025-89124-8

He, Y., Hu, J., Zeng, M., Qian, Y., & Zhang, R. (2024). DCGC-YOLO: The Efficient Dual-Channel Bottleneck Structure YOLO Detection Algorithm for Fire Detection. *IEEE Access*, *12*(April), 65254–65265. https://doi.org/10.1109/ACCESS.2024.3385856

Hussain, M. (2024). YOLOv1 to v8: Unveiling Each Variant-A Comprehensive Review of YOLO. *IEEE Access*, *12*(February), 42816–42833. https://doi.org/10.1109/ACCESS.2024.3378568

Khan, A. (2024). The Emergence of the Fourth Industrial Revolution and its Impact on International Trade. *ASR: Chiang Mai University Journal of Social Sciences and Humanities*, *11*(1), 1–13. https://doi.org/10.12982/cmujasr.2024.007

Malahina, E. A. U., Saitakela, M., Bulan, S. J., Lamabelawa, M. I. J., & Belutowe, Y. S. (2024). Teachable Machine: Optimization of Herbal Plant Image Classification Based on Epoch Value, Batch Size and Learning Rate. *Journal of Applied Data Sciences*, *5*(2), 532–545. https://doi.org/10.47738/jads.v5i2.206

Onososen, A. O., Musonda, I., Onatayo, D., Saka, A. B., Adekunle, S. A., & Onatayo, E. (2025). Drowsiness Detection of Construction Workers: Accident Prevention Leveraging Yolov8 Deep Learning and Computer Vision Techniques. *Buildings*, *15*(3). https://doi.org/10.3390/buildings15030500

Prince, R. H., Mamun, A. Al, Peyal, H. I., Miraz, S., Nahiduzzaman, M., Khandakar, A., & Ayari, M. A. (2024). CSXAI: a lightweight 2D CNN-SVM model for detection and classification of various crop diseases with

Literasi Nusantara. vol.5 no. 3, July 2025 – October 2025
ISSN 2746 8208 (Online) and ISSN 2746 3575 (Print)

explainable AI visualization. *Frontiers in Plant Science*, *15*(July), 1–18. https://doi.org/10.3389/fpls.2024.1412988

Sahafi, A., Koulaouzidis, A., & Lalinia, M. (2024). Polypoid Lesion Segmentation Using YOLO-V8 Network in Wireless Video Capsule Endoscopy Images. *Diagnostics*, *14*(5), 1–16. https://doi.org/10.3390/diagnostics14050474

Sun, W., Ren, S., & Tang, G. (2025). In the era of responsible artificial intelligence and digitalization: business group digitalization, operations and subsidiary performance. *Annals of Operations Research*. https://doi.org/10.1007/s10479-024-06453-z

Tang, X. X., & Guo, P. Y. (2024). Fatigue Driving Detection Methods Based on Drivers Wearing Sunglasses. *IEEE Access*, *12*(April), 70946–70962. https://doi.org/10.1109/ACCESS.2024.3394218

Vijayakumar, A., & Vairavasundaram, S. (2024). YOLO-based Object Detection Models: A Review and its Applications. In *Multimedia Tools and Applications* (Issue 0123456789). Springer US. https://doi.org/10.1007/s11042-024-18872-y

Widianto, F. (2023). Examining Characteristics on Twitter Users' Text and Hashtag Utilization During Tech Winter Layoff Post-COVID-19 Using LDA and K-Means Clustering Approach. *Makara Human Behavior Studies in Asia*, *27*(2). https://scholarhub.ui.ac.id/hubsasia.https://doi.org/10.7454/hubs.asia.1191223