AI-Based Performance Analysis of YOLOv8 for Computer User Health Monitoring: Posture, Drowsiness, and Eye Gaze Detection

Aeron Red R. Celajes
New Era University
aeronred.celajes@neu.edu.ph

Angelo Gerard T. Mallari
New Era University
angelogerard.mallari@neu.edu.ph

ABSTRACT

The proposed paper aims to develop and analyze a multiple AI solution model using YOLOv8 for computer user health monitoring, that only focuses on improper sitting postures, signs of drowsiness, and disengagement, where eye gaze is used to detect it, during computer usage. The researchers decided to create 3 models and combine it as one for deployment.

The researchers focused on collecting and annotating the dataset, training the model, and evaluating its performance. It was tested and trained using Roboflow where the researchers used publicly available datasets to customize it, and the measurement of performance of the model was measured in terms of mean Average Precision (mAP), Precision, and Recall.

There are three YOLOv8 models, for posture it achieved a high mAP of 78.8%, a Precision of 70.1%, and a Recall of 75.5%. Then for drowsiness it achieved a high mAP of 97.2%, a Precision of 95.4%, and a Recall of 95.3%. Lastly for eye gaze, it achieved a high mAP of 95.0%, a Precision of 92.7%, and a Recall of 95.6%. The conclusion of this study is that the YOLOv8 model is indeed effective for the three models when it comes for computer user health monitoring, hence rejecting the null hypothesis, but for the posture model it could have some improvements especially the datasets to make the data better.

Keywords:
YOLOv8, Health Monitoring, Posture Detection, Drowsiness Detection, Eye Gaze Detection, mAP, Precision, Recall, Roboflow

1. INTRODUCTION

In today's age, more individuals rely on computers for everyday life. Using computers has multiple advantages, but computers also lead to serious health issues related to continuous computer use. Such habits often lead to improper posture, which is one related computer health issue, and this improper posture can cause disturbances in the musculoskeletal balance and may disturb physiological processes other than the musculoskeletal system [5].

Health surveillance is a core function for the systematic collection, analysis, and interpretation of outcome data for the purpose of evaluating and improving health practices [10]. There are times that individuals tend to not notice their own health behavior, wherein computer user health monitoring can be beneficial for treating health related concerns.

This project aims to generate a comprehensive dataset, train a YOLOv8 model for precise detection, and then examine the visual characteristics of the different health issues for computer users such as poor postures, drowsiness, and eye gaze. This project works on these aspects by creating a health-monitoring AI model that uses computer vision.

1.1 Objectives

This project aims to achieve the following:

1. To develop a model that uses YOLOv8 to monitor computer users' posture and detect unhealthy sitting positions.
2. To develop a model that uses YOLOv8 to identify signs of drowsiness.
3. To develop a model that uses YOLOv8 to analyze eye gaze patterns, when the eyes are close or not.
4. To evaluate the effectiveness of the model by its performance using mAP, Precision, and Recall.

1.2 Research Gap

Although object detection models are widely used, using an AI-multiple solution model in health monitoring for computer users is still limited. Usually, studies for computer user health monitoring that use computer vision solely focus on one health issue, like only focusing on posture monitoring wherein one recent study stated that the objective of the research is to design and implement e Internet of Things (IoT) devices and Google's Media Pipe library for detecting bad posture [14]. This project seeks to address this gap by providing multiple AI solutions for computer user health by focusing on different health-related issues such as posture, drowsiness, and disengagement, where eye gaze is used to detect it, where publicly available datasets are customized by the researchers and evaluating the performance of YOLOv8. This is tested in Roboflow for evaluation of the model.

1.3 Hypothesis

YOLOv8 can effectively detect posture, drowsiness, and eye gaze in computer users with high mAP, precision, and recall, which makes it a reliable model for computer user health monitoring.

2. LITERATURE REVIEW / STUDIES

Computer vision has become important in detecting and classifying objects in real-time. YOLO (You Only Look Once), introduced by Redmon et al., develops a real-time object detection with an outstanding balance of speed and accuracy [12]. A recent model, YOLOv8, just enhances these qualities even more, making it an excellent choice for applications that need real-time detection with high precision.

2.1 Previous Applications of YOLO Models

YOLO models are very effective in many fields of application. For example, YOLOv3 was successful in demonstrating the detection of fruits and other agricultural products with impressive accuracy and speed. [8]. YOLOv4 was introduced afterwards wherein its performance in terms of accuracy and speed was better when it was put into use as a detector that identifies face masks, which will be effective for public health monitoring [16]. Then, YOLOv5 further improved, giving greater speed and accuracy than YOLOv4 and YOLOv3, which just continues to improve the model itself as time goes by [17]. For YOLOv8, one recent study stated that this model was used for medical purposes, which was done for brain tumor detection and medical imaging [11]. This indicates that YOLOv8 has high potential to assist in high risk situations such as being used in the medical field where it shows its high ability for object detection.

2.2 Use in Health Monitoring and Ergonomics

The monitoring of health has changed over time since technology continuously changes and improves. Traditional techniques have been highly based on the self-reported questionnaire or observational study. This kind of approach has often suffered bias from recall and measurement errors [6].

Recent approaches have been the use of wearable devices, wherein it gives improved accurate, and objective ways of measuring physical activity and sedentary behavior. These kinds of approaches are widely accepted as it is an emerging approach and it is really excellent for improving human activities and quality of life [15].

Another approach is by using computer vision wherein a recent study presented on detecting upper body sedentary behaviors with cameras to capture images of the user's posture, which are then analyzed to detect any improper posture of the upper body [7].

2.3 Comparative Analysis of Object Detection Models

Compared to other object detection models like Faster R-CNN and YOLO models, it was compared and results in YOLO being better in performance making YOLO ideal for real-time applications [1]. Afterwards, YOLOv8 improves on its previous models when it comes to balancing of speed and accuracy [9].

2.4 Challenges and Future Directions

Object detection is widely used in many applications, including autonomous driving and surveillance to robotics and visual search engines. The challenges that can occur in object detection are such as customizing the dataset itself considering the variation of pixels, processing of low-resolutions, handling different sizes of multiple objects, the availability of the labeled data, and handling overlapping objects [4]. This depicts that most machine learning and deep learning based models require a very large set of data to be used for training and annotation. The performance lowers if there is less data.

3. METHODOLOGIES

3.1 Dataset Collection

The researchers gathered datasets for this study from publicly available sources such as Roboflow, which include annotated images on posture detection, drowsiness monitoring, and eye gaze detection. Additionally, the researchers tested the model in real-time to assess the model's performance.

3.2 Data Preprocessing

The collected images for training were resized to 640x640 pixels and were auto-orientated. This is to ensure an optimal model performance.

3.3 Model Training

The researchers chose the YOLOv8 model, as this model is excellent for object detection and classification. The Roboflow platform was used to enhance the training process of the data as it offers data preprocessing model training and evaluation. The model was trained in Google Colab, utilizing a T4 GPU to accelerate computations, but adjustments were made to hyperparameters such as the learning rate, batch size, and the number of epochs. This is to reduce overfitting and ensure high accuracy.

3.4 Architecture of the YOLOv8 Model

The architecture of YOLOv8 was designed to balance speed and accuracy in object detection. This model uses Cross Stage Partial (CSP) as its backbone for efficient feature extraction, and its neck uses a Feature Pyramid Network (FPN) and Path Aggregation Network (PAN) to enhance multi-scale feature fusion, and its head introduces a decoupled head architecture which separates the classification and detection processes [13]. Because of its architecture, YOLOv8 is an excellent choice for posture, drowsiness, and eye gaze, as it can use bounding boxes, forecast class probabilities, and show the objectness score all at once.

3.5 Annotations

The dataset which was from Roboflow included TXT files containing annotations. These were used to train the model, which reduces the manual effort for labeling. This is to ensure high accuracy and reliability in categorizing the images.

3.6 Evaluation Metrics

The performance of the YOLOv8 model was evaluated using standard metrics for object detection. This includes Precision, Recall, and mAP/mAP50 which will help in determining the assessment of the model's performance. Additionally, mAP50-95, Confusion Matrix, F1-Score, Box loss, Classification loss, and Distribution Focal loss will offer deeper insights into the model's classification performance.

3.7 Data Splitting and Cross-Validation

For the Posture Detection Model has a total of 2290 images, the train set has 74% or 1690 images, for the valid set it contains 500 images or 22%, and the test set has a total of 100 images or 4%. For the Drowsiness

Detection Model, it has a total of 1524 images divided into 3 sets, the train set contains 1524 images or 87%, the valid set with 8% or 139 images, and the test set which has 91 images or 5%. Lastly, the Eye Gaze Detection Model, there is a total of 1243 images with the help of data augmentation, which has been divided into 3 sets, train set which has 1185 images or 95%, the valid set which has 41 images or 3%, and the test set which has 17 images or 1%.

The Stratified splitting ensures that all 5 classes (good posture, bad posture, drowsy, eyes closed, and eyes open) are equally represented in each set. It lessens the chance of overfitting and makes it possible to evaluate the model's performance in detail. Cross-validation techniques may be studied in the future to verify the model's reliability across different data splits.

4. RESULTS AND DISCUSSIONS

Now, the performance of the YOLOv8 model was evaluated using various evaluation metrics. This demonstrates its effectiveness in accurately detecting and classifying health-related behaviors such as posture, drowsiness, and eye gaze. The following section provides a detailed analysis of the results based on these metrics.

4.1 Posture Model Metrics

4.1.1 F1 - Confidence Curve for Posture

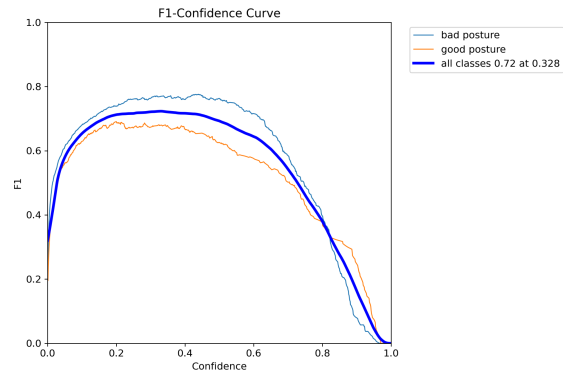

Figure 1: F1 - Confidence Curve for Posture

In Figure 1, the relationship between the F1 score and confidence levels are shown for each posture class. Now the overall score for F1 score for all classes is 0.72, while the confidence level is 0.328. This means that the F1 score is still decent and it demonstrates its moderate effectiveness in accurately identifying the posture classes across different confidence thresholds.

4.1.2 Confusion Matrix for Posture

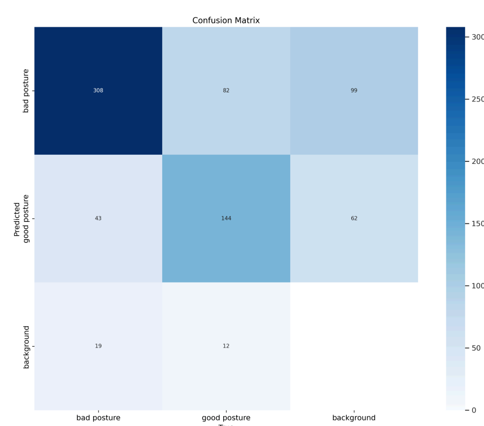

Figure 2: Confusion Matrix for Posture

Figure 2 shows the graph of the model's classification performance for each of the posture classes. The model correctly detects 308 correct predictions for bad posture, indicating high accuracy for it. Now for good posture, it correctly detects 144, with decent accuracy, although there are some misclassifications. This indicates that even though the model has high accuracy, improvements could still be done to reduce misclassification rates.

4.1.3 mAP50 for Posture

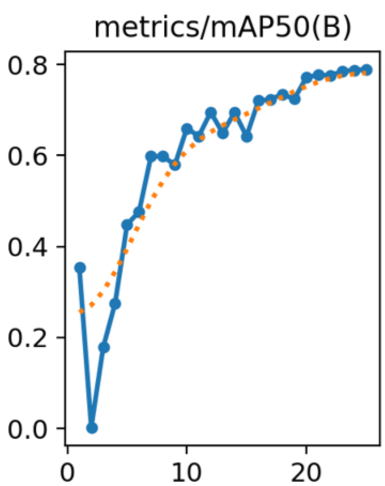

Figure 3: mAP50 for Posture

In Figure 3, this shows the mean Average Precision at 50% details. In the first epoch, mAP50 was 0.35424, and then it steadily increased up until the last epoch, which is epoch 25, it reached 0.788, but the optimized mAP50 is 0.78798 in epoch 24. This improvement indicates the model’s enhanced ability to accurately localize and classify objects with moderate performance sustaining within the final epochs.

4.1.4 mAP50-95 for Posture

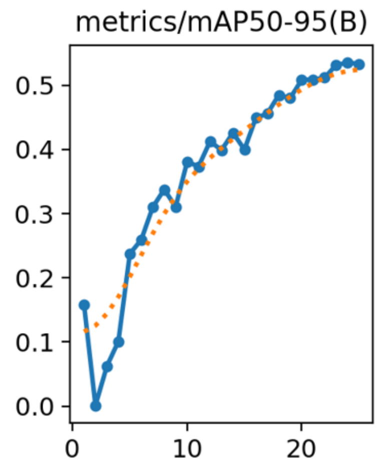

Figure 4: mAP50-95 for Posture

In Figure 4, this shows the mean Average Precision at 50% to 95% details. In the first epoch, mAP50 was 0.15866, and then it steadily increased up until the last epoch, which is epoch 25, it reached 0.5334. This improvement still shows the model’s performance in detecting objects under different levels of overlap, especially in later epochs.

4.1.5 Precision for Posture

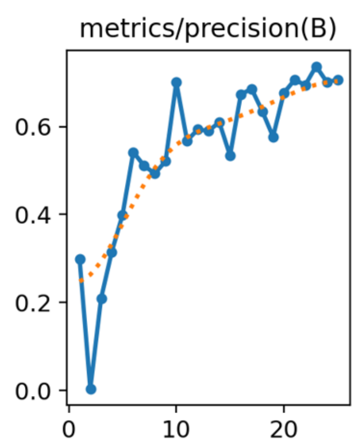

Figure 5: Precision for Posture

Figure 5 illustrates the precision of the model. The precision at epoch 1 starts at 0.29952 but at epoch 25 it ends with 0.70614, but the optimized precision is 0.70137 in epoch 24. This indicates that as the training progresses, the model becomes better at identifying correct posture classes, but it could have an improvement to have a higher precision overall.

4.1.6 Recall for Posture

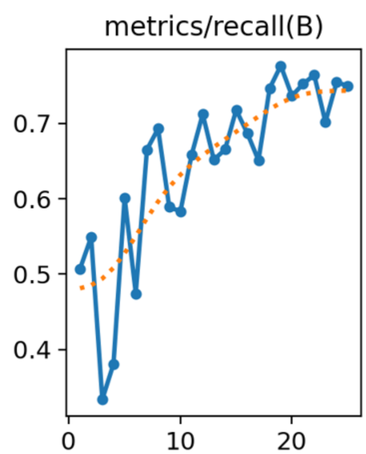

Figure 6: Recall for Posture

Figure 6 shows the recall of the model. The recall at the beginning reaches 0.50708 at epoch 1, then it reaches 0.7502 at epoch 25, but the optimized recall is 0.75461 in epoch 24. This implies that the model improves at capturing relevant instances as training goes on, but improvements can be done as it may still be able to achieve a higher overall recall.

4.1.7 Box Loss for Posture

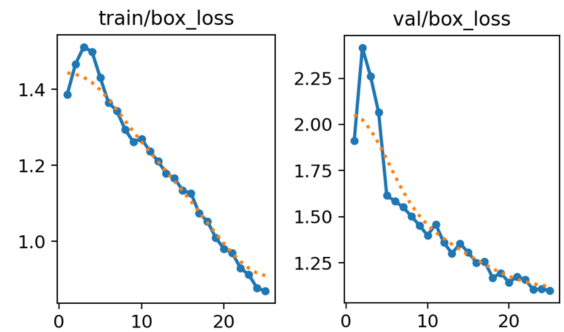

Figure 7: Box Loss for Posture

In Figure 7, it shows the box loss for posture during training and validation. Here, the training box loss at epoch 1 started with 1.3869 and then ended with 0.86893 at epoch 25, which indicates a significant improvement of the model to accurately predict the bounding boxes, since at first the model struggled but after until the last epoch, the model improved since the box loss was lowered.

4.1.8 Classification Loss for Posture

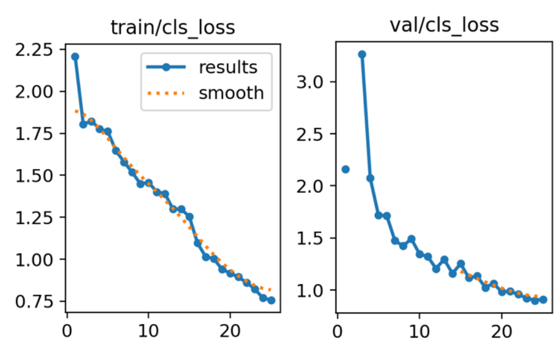

Figure 8: Classification Loss for Posture

In Figure 8, the classification loss for posture during training and validation is displayed. In this case, the training box loss at epoch 1 began at 2.2093 and ended at 0.75621 at epoch 25, indicating a notable improvement in the model's ability to accurately learn the fundamentals of classification. Initially, the model struggled, but as the classification loss decreased until the final epoch, the model improved.

4.1.9 Distribution Focal Loss for Posture

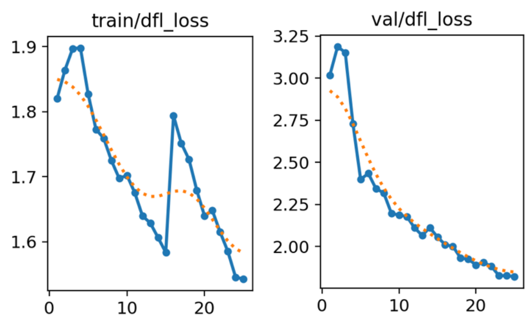

Figure 9: Distribution Focal Loss for Posture

In Figure 9, The distribution focal loss for posture during training and validation is displayed. In this case, the training box loss at epoch 1 began at 1.8205 and ended at epoch 25 which is 1.5426. This shows that the model has significantly improved to concentrate on more difficult-to-classify examples, which is helpful for imbalanced datasets. Initially, the model struggled, but as the distribution focal loss decreased until the last epoch, the model improved.

4.2 Drowsiness Model Metrics

4.2.1 F1 - Confidence Curve for Drowsiness

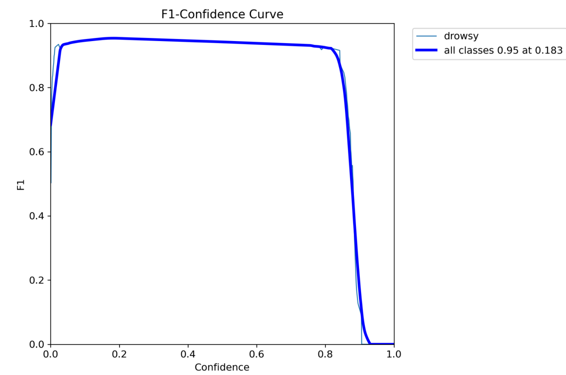

Figure 10: F1 - Confidence Curve for Drowsiness

In Figure 10, the relationship between the F1 score and confidence levels are shown for the drowsy class. Now the overall score for F1 score for the drowsy class is 0.95, while the confidence level is 0.183. This means that the F1 score is high and it demonstrates its effectiveness in accurately identifying the drowsy class across different confidence thresholds.

4.2.2 Confusion Matrix for Drowsiness

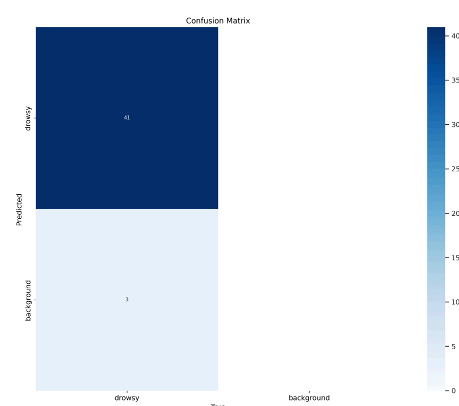

Figure 11: Confusion Matrix for Drowsiness

Figure 11 shows the graph of the model's classification performance for the drowsy class. The model correctly detects 41 correct predictions for drowsy, indicating high accuracy for it although there are some misclassifications. This indicates that even though the model has high accuracy, improvements could still be done to reduce misclassification rates.

4.2.3 mAP50 for Drowsiness

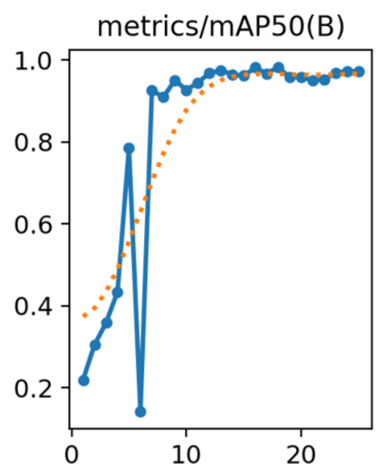

Figure 12: mAP50 for Drowsiness

In Figure 12, this shows the mean Average Precision at 50% details. In the first epoch, mAP50 was 0.21902, and then it steadily increased up until the last epoch, which is epoch 25, it reached 0.97229. This improvement indicates the model's enhanced ability to accurately localize and classify objects with high performance sustaining within the final epochs.

4.2.4 mAP50-95 for Drowsiness

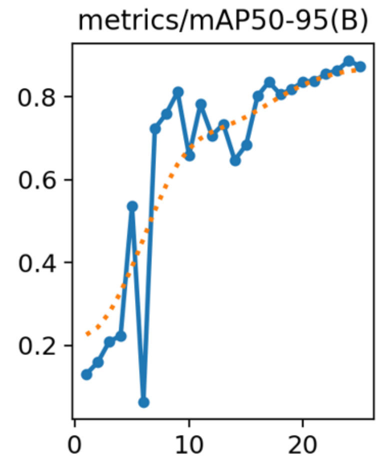

Figure 13: mAP50-95 for Drowsiness

In Figure 13, this shows the mean Average Precision at 50% to 95% details. In the first epoch, mAP50 was 0.13144, and then it steadily increased up until the last epoch, which is epoch 25, it reached 0.87356. This improvement still shows the model's performance in detecting objects under different levels of overlap, especially in later epochs.

4.2.5 Precision for Drowsiness

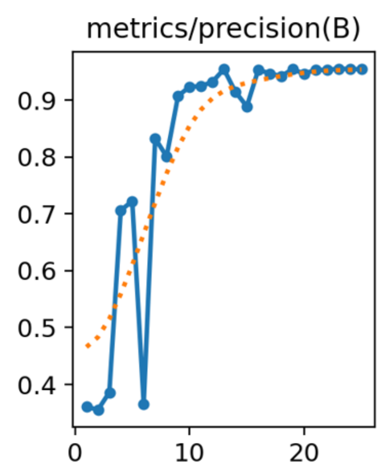

Figure 14: Precision for Drowsiness

Figure 14 illustrates the precision of the model. The precision at epoch 1 starts at 0.36134 but at epoch 25 it ends with 0.95448. This indicates that as the training progresses, the model becomes better at identifying the correct drowsy class, wherein this model is highly effective at it.

4.2.6 Recall for Drowsiness

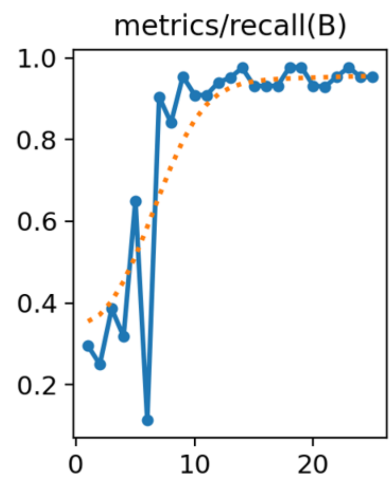

Figure 15: Recall for Drowsiness

Figure 15 shows the recall of the model. The recall at the beginning reaches 0.29545 at epoch 1, then it reaches 0.9532 at epoch 25. This implies that the model has a high effective ability in capturing relevant instances as it even improves as training goes on.

4.2.7 Box Loss for Drowsiness

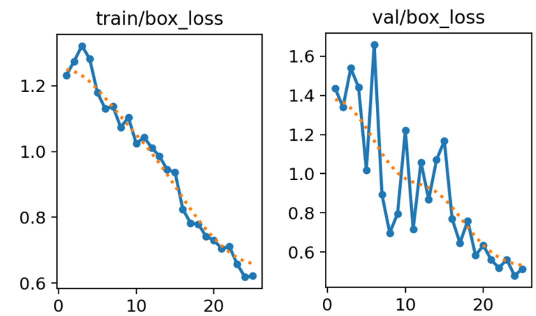

Figure 16: Box Loss for Drowsiness

In Figure 16, it shows the box loss for drowsiness during training and validation. Here, the training box loss at epoch 1 started with 1.2325 and then ended with 0.62114 at epoch 25, which indicates a significant improvement of the model to accurately predict the bounding boxes, since at first the model struggled but after until the last epoch, the model improved since the box loss was lowered.

4.2.8 Classification Loss for Drowsiness

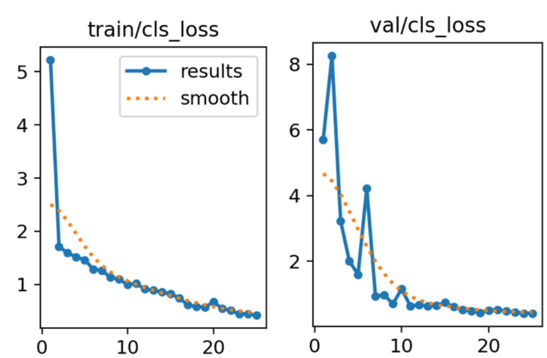

Figure 17: Classification Loss for Drowsiness

In Figure 17, the classification loss for drowsiness during training and validation is displayed. In this case, the training box loss at epoch 1 began at 5.2291 and ended at

0.41699 at epoch 25, indicating a notable improvement in the model's ability to accurately learn the fundamentals of classification. Initially, the model struggled, but as the classification loss decreased until the final epoch, the model improved.

4.2.9 Distribution Focal Loss for Drowsiness

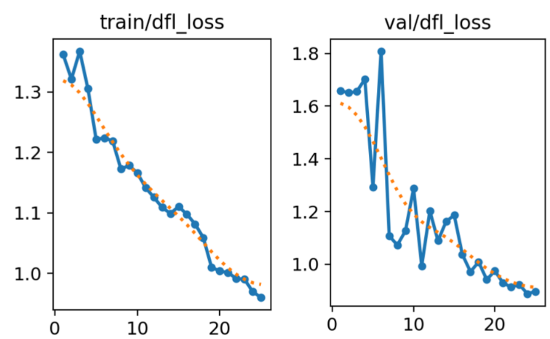

Figure 18: Distribution Focal Loss for Drowsiness

In Figure 18, The distribution focal loss for drowsiness during training and validation is displayed. In this case, the training box loss at epoch 1 began at 1.362 and ended at epoch 25 at 0.95997. This shows that the model has significantly improved to concentrate on more difficult-to-classify examples, which is helpful for imbalanced datasets. Initially, the model struggled, but as the distribution focal loss decreased until the last epoch, the model improved.

4.3 Eye Gaze Model Metrics

4.3.1 F1 - Confidence Curve for Eye Gaze

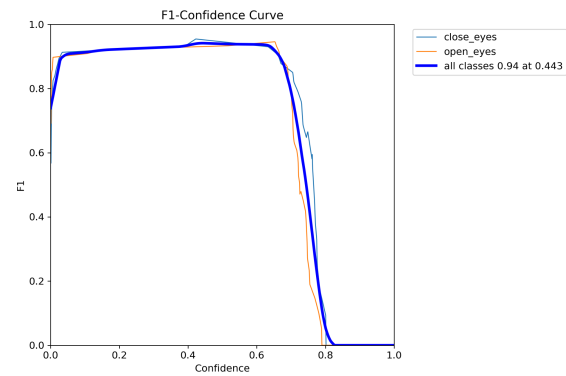

Figure 19: F1 - Confidence Curve for Eye Gaze

In Figure 19, the relationship between the F1 score and confidence levels are shown for each eye gaze class. Now the overall score for F1 score for all classes is 0.94, while the confidence level is 0.443. This means that the F1 score is high and it demonstrates its effectiveness in accurately identifying the eye gaze classes across different confidence thresholds.

4.3.2 Confusion Matrix for Eye Gaze

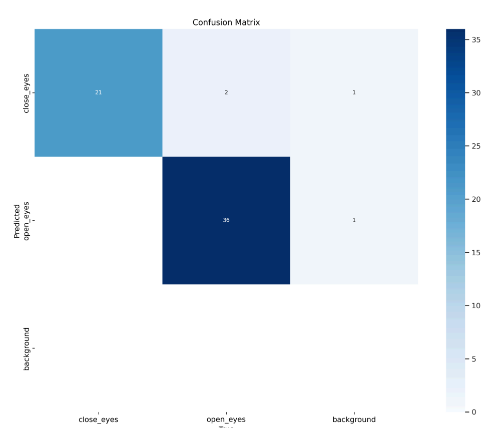

Figure 20: Confusion Matrix for Eye Gaze

Figure 20 shows the graph of the model's classification performance for each of the eye gaze classes. The model correctly detects 36 correct predictions for open eyes, indicating high accuracy for it. While for close eyes, it correctly detects 21, although there are some misclassifications. This indicates that even though the model has high accuracy, improvements could still be done to reduce misclassification rates.

4.3.3 mAP50 for Eye Gaze

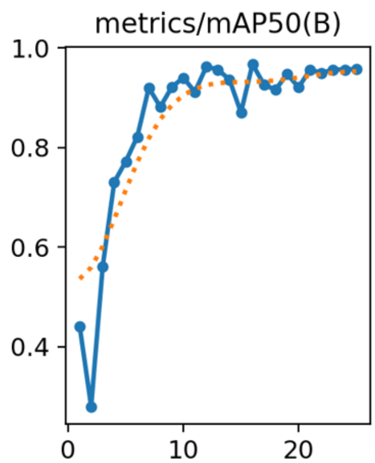

Figure 21: mAP50 for Eye Gaze

In Figure 21, this shows the mean Average Precision at 50% details. In the first epoch, mAP50 was 0.44136, and then it steadily increased up until the last epoch, which is epoch 25, it reached 0.95751, but the optimized mAP50 is 0.9496 in epoch 22. This improvement indicates the model's enhanced ability to accurately localize and classify objects with high performance sustaining within the final epochs.

4.3.4 mAP50-95 for Eye Gaze

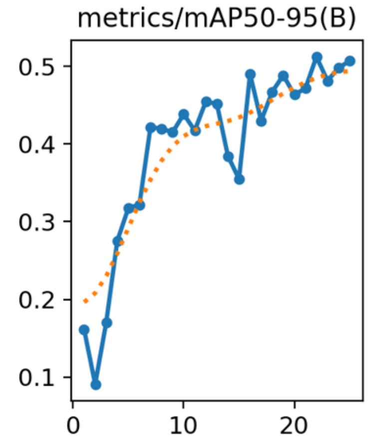

Figure 22: mAP50-95 for Eye Gaze

In Figure 22, this shows the mean Average Precision at 50% to 95% details. In the first epoch, mAP50 was 0.16127, and then it steadily increased up until the last epoch, which is epoch 25, it reached 0.50692. This improvement still shows the model's performance in detecting objects under different levels of overlap, especially in later epochs.

4.3.5 Precision for Eye Gaze

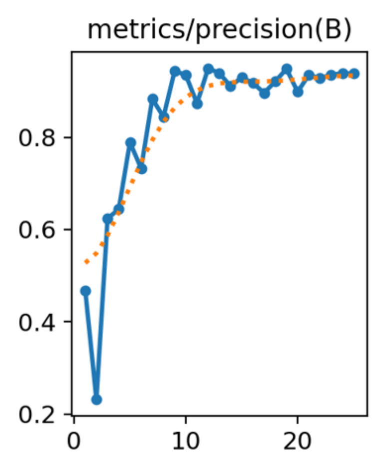

Figure 23: Precision for Eye Gaze

Figure 23 shows the precision of the model. The precision at epoch 1 starts at 0.46839 but at epoch 25 it ends with 0.93768, but the optimized precision is 0.9274 in epoch 22. This depicts that as training goes on, the model improves its ability to recognize the appropriate eye gaze class, which makes the model effective in doing the task.

4.3.6 Recall for Eye Gaze

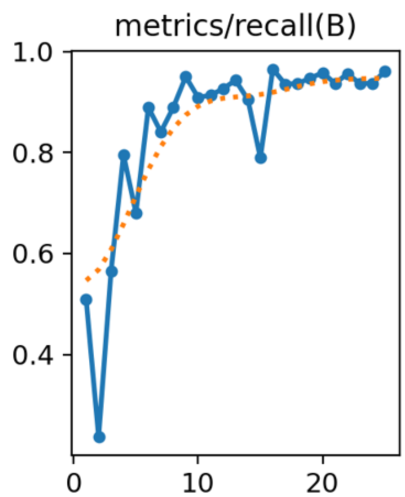

Figure 24: Recall for Eye Gaze

Figure 24 displays the model's recall. At epoch 1, the recall reaches 0.50877, and at epoch 25, it reaches 0.96053, but the optimized recall is 0.95586 in epoch 22. Given that it even gets better with training, this suggests that the model has a high effective ability to capture relevant instances.

4.3.7 Box Loss for Eye Gaze

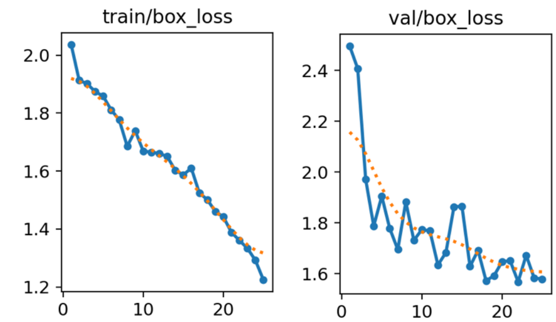

Figure 25: Box Loss for Eye Gaze

In Figure 25, it shows the box loss for eye gaze during training and validation. Here, the training box loss at epoch 1 started with 2.0369 and then ended with 1.2245 at epoch 25, which indicates a significant improvement of the model to accurately predict the bounding boxes, since at first the model struggled but after until the last epoch, the model improved since the box loss was lowered.

4.3.8 Classification Loss for Eye Gaze

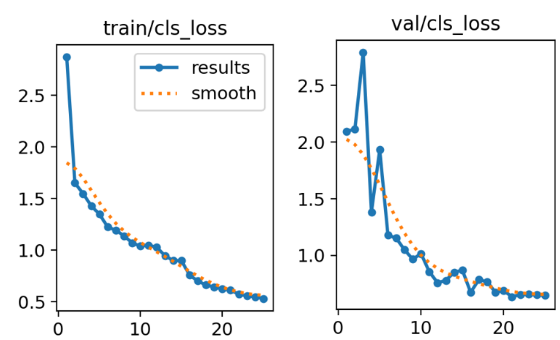

Figure 26: Classification Loss for Eye Gaze

In Figure 26, the classification loss for eye gaze during training and validation is displayed. In this case, the training box loss at epoch 1 began at 2.8736 and ended at 0.52817 at epoch 25, indicating a notable improvement in the model's ability to accurately learn the fundamentals of classification. Initially, the model struggled, but as the classification loss decreased until the final epoch, the model improved.

4.3.9 Distribution Focal Loss for Eye Gaze

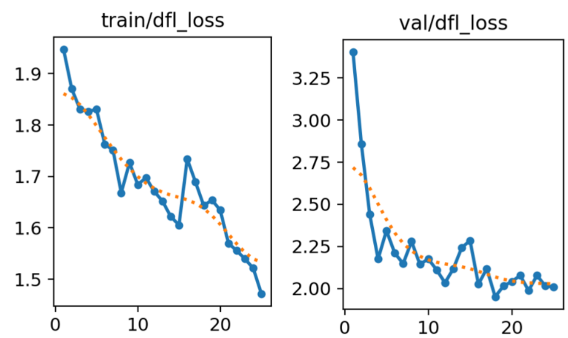

Figure 27: Distribution Focal Loss for Eye Gaze

In Figure 27, The distribution focal loss for eye gaze during training and validation is displayed. In this case, the training box loss at epoch 1 began at 1.9478 and ended at epoch 25 at 1.4713. This shows that the model has significantly improved to concentrate on more difficult-to-classify examples, which is helpful for imbalanced datasets. Initially, the model struggled, but as the distribution focal loss decreased until the last epoch, the model improved.

4.4 Sample Output of Object Detection on

The images below show the output generated by the YOLOv8 model in detecting and classifying various health-related behaviors for computer users specifically for posture, drowsiness and eye gaze. This demonstration aims to give a clear illustration of how the model functions in recognizing and labeling various computer components in an actual situation. This sample output highlights the model's ability to generalize and precisely identify computer components outside of the training data by utilizing an independently gathered dataset.

Posture Detection

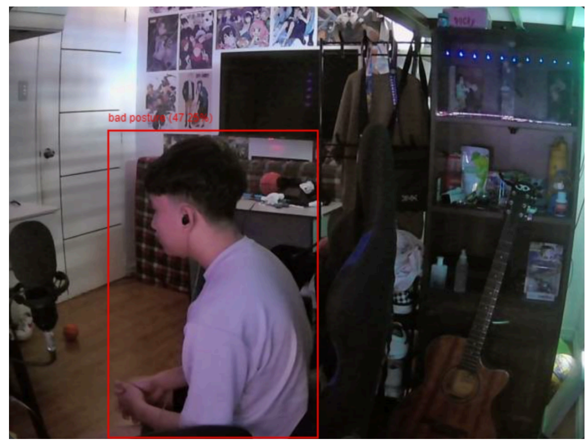

Figure 28: Bad Posture

Figure 28 shows the YOLOv8 successfully detects and classifies the bad posture with a confidence level of 47.25%. The bad posture is highlighted with a bounding box and labeled “bad posture” along with the confidence percentage.

Figure 29: Good Posture

Figure 29 shows that the YOLOv8 can successfully recognize and classify good posture with a 60% confidence level. The labeled "good posture" is shown along with the confidence percentage, and the good posture is highlighted with a bounding box.

Drowsiness Detection

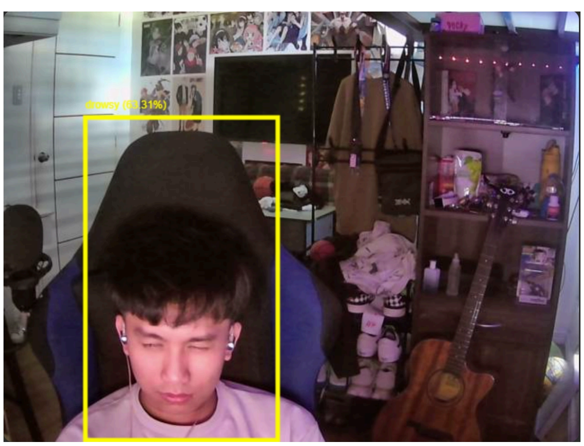

Figure 30: Drowsy

In Figure 30, the YOLOv8 has a 63.31% confidence level in its ability to recognize and classify drowsiness. A bounding box is used to highlight the drowsy, and "drowsy" is labeled with the confidence percentage.

Eye Gaze Detection

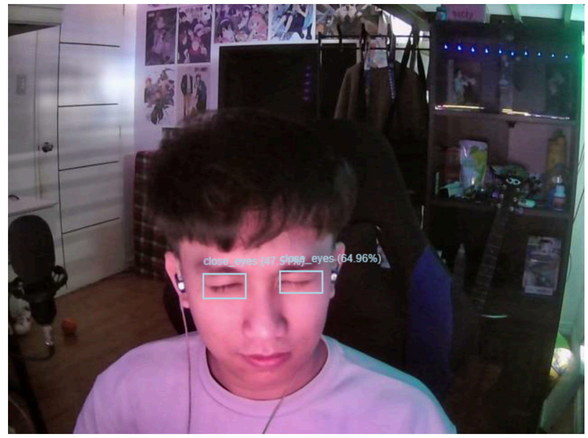

Figure 31: Close Eyes

In Figure 31, the YOLOv8's confidence level in identifying and categorizing close eyes is 64.96%. The "close_eyes" is labeled with the confidence percentage, and the close eyes are highlighted with a bounding box.

Figure 32: Open Eyes

In Figure 32, the YOLOv8 has a 60.75% confidence level in recognizing and classifying open eyes. A bounding box is used to highlight open eyes, and "open_eyes" is labeled with the confidence percentage.

Multiple Health-related Behaviors

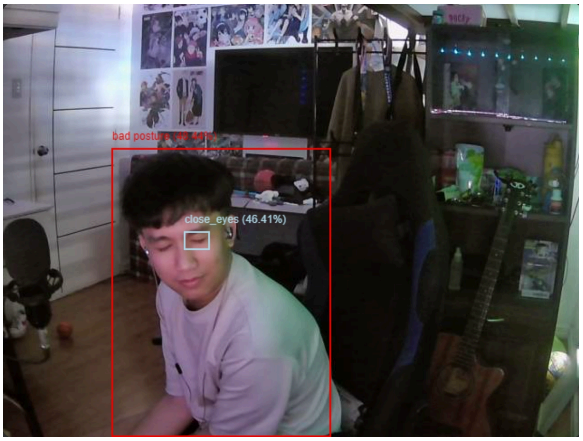

Figure 33: Bad Posture and Close Eyes

In Figure 33, the YOLOv8 has a 48.44% confidence level in recognizing and classifying bad posture. And it also recognizes the close eyes with a confidence level of 46.41%. This showcases the ability to detect multiple health-related issues at once.

5. CONCLUSION

5.1 Conclusion

This research aimed to develop and evaluate an object detection model using the YOLOv8 architecture to accurately identify and classify health related issues of computer users, particularly only the posture, drowsiness and disengagement, where eye gaze is used to detect it.

Basically the researchers used 3 models and combined them into one for deployment using Roboflow. For the posture model, there are a total of 2290 datasets which were divided into two classes which are: good posture and bad posture. For the drowsiness model, it has 1524 datasets and it has only one class which is the drowsy class. Lastly for the eye gaze model, it has a total of 1243 datasets which is divided into two classes which are: close eyes and open eyes.

Throughout the training process, the researchers used the three key metrics for evaluating the model's performance, which consists of mAP, precision and recall. For posture it achieved a high mAP of 78.8%, a Precision of 70.1%, and a Recall of 75.5%. Then for drowsiness it achieved a high mAP of 97.2%, a Precision of 95.4%, and a Recall of 95.3%. Lastly for eye gaze, it achieved a high mAP of 95.0%, a Precision of 92.7%, and a Recall of 95.6%.

The study concludes that the YOLOv8 model is effective for computer user health monitoring when compared to the other three models, thereby rejecting the null hypothesis. However, the posture model could benefit from some improvements, particularly in terms of datasets to improve the data. This study can help in assisting in the health of computer users as it achieved a strong mean average precision, precision, and recall. Future research can focus on optimizing the model's architecture, computational efficiency, and the datasets.

5.2 Recommendation

Future researchers should explore integrating audio cues and by adding these features, like alerts and warnings for example to the user if the system detects something that is unhealthy for them, it would help them to learn to keep their posture and health in check and to develop good habits. By expanding datasets with diverse images and classes, it will improve the quality of the model and its detection to further give higher and more accurate results.

5.3 Limitations of the Study

The study's performance is limited by the availability of labeled datasets and the model's sensitivity to lighting, video quality, and surroundings seen in the video feed. Due to the limited supply of datasets that are available on the internet, there are not a lot of datasets that shows the user's posture from upfront, most of them are taken from the side of the user, because of that, it limits the detection of the user's posture to only it's side and not the frontal posture. Another limitation is the number epochs and batch size for training the YOLOv8 model, due to Google Colab's free feature that is used in this study, it cannot handle training a large amount of epochs and batch size so it limits the probability of getting a higher accuracy for the models. Roboflows data augmentation has its limits too, if you do not have the premium version, you can only multiply the chosen dataset up to three times, this limits the dataset used for training that could potentially provide better results.

References

[1] Abdulghani, A. M. A., & Menekşe Dalveren, G. G. (2022). Moving Object Detection in Video with Algorithms YOLO and Faster R-CNN in Different Conditions. *Avrupa Bilim Ve Teknoloji Dergisi(33)*, 40-54. https://doi.org/10.31590/ejosat.1013049

[2] Armecin Workspace. (n.d.). improper-sitting Object Detection Dataset. *Roboflow*. https://universe.roboflow.com/armecin-workspace/improper-sitting-av8w4

[3] Augmented Startups (2022, December). Drowsiness detection object detection dataset and Pre-Trained model by augmented startups. *Roboflow*. https://universe.roboflow.com/augmented-startups/drowsiness-detection-cntmz

[4] Diwan, T., Anirudh, G. & Tembhurne, J.V. (2023). Object detection using YOLO: challenges, architectural successors, datasets and applications. *Multimed Tools Appl* 82, 9243–9275 . https://doi.org/10.1007/s11042-022-13644-y

[5] Fathima, D., Lobo, J., Angioi, M., Blach, W., Rydzik, L., Ambroży, T. & Malliaropoulos, N. (2024, August 09). Sedentary Lifestyle, Heart Rate Variability, and the Influence on Spine Posture in Adults: A Systematic Review Study. *Appl. Sci. 2024*, 14(16), 6985. https://doi.org/10.3390/app14166985

[6] Gao, Z., Liu, W., McDonough, D. J., Zeng, N., & Lee, J. E. (2021). The Dilemma of Analyzing Physical Activity and

Sedentary Behavior with Wrist Accelerometer Data: Challenges and Opportunities. *Journal of Clinical Medicine*, 10(24), 5951. https://doi.org/10.3390/jcm10245951

[7] Guduru, R. K. R., Domeika, A., Dubosiene, M., & Kazlauskiene, K. (2021). Prediction framework for upper body sedentary working behaviour by using deep learning and machine learning techniques. *Soft Computing*, 26, 12969–12984. https://doi.org/10.1007/s00500-021-06156-8

[8] Lawal, M.O. Tomato detection based on modified YOLOv3 framework. *Sci Rep 11*, 1447 (2021). https://doi.org/10.1038/s41598-021-81216-5

[9] Mohan, M., Ali, L., Muhammed Swavaf, Salah Bouktif, Munkhjargal Gochoo, Hamad Aljassmi, & Alnajjar, F. (2024). Navigating the YOLO Landscape: A Comparative Study of Object Detection Models for Emotion Recognition. *IEEE Access*, 12, 109427–109442. https://doi.org/10.1109/access.2024.3439346

[10] Omura, J. D., Whitfield, G. P., Chen, T. J., Hyde, E. T., Ussery, E. N., Watson, K. B., & Carlson, S. A. (2021). Surveillance of Physical Activity and Sedentary Behavior Among Youth and Adults in the United States: History and Opportunities. *Journal of Physical Activity and Health*, 18(S1), S6-S24. https://doi.org/10.1123/jpah.2021-0179

[11] Patel, S., Kadam, Y., Thombare, A., Salvi, K., Jadhav, D. (2024). A review of brain tumor detection techniques using yolov8. *International Journal for Research in Applied Science and Engineering Technology*, 12(3), 1075-1078. https://doi.org/10.22214/ijraset.2024.58666

[12] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You Only Look Once: Unified, RealTime Object Detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). https://doi.org/10.48550/arXiv.1506.02640

[13] Solimani, F., Cardellicchio, A., Dimauro, G., Petrozza, A., Summerer, S., Cellini, F., & Renò, V. (2024, March). Optimizing tomato plant phenotyping detection: Boosting YOLOv8 architecture to tackle data complexity. *Computers and Electronics in Agriculture*, 218, 108728. https://doi.org/10.1016/j.compag.2024.108728

[14] Sreevani, V., Reddy, B. S., Nithin, K., Vardhan, K. H., Mohammed, K. A., Reddy, U., Lakhanpal, S., & Kalra, R. (2024). Advanced interdisciplinary approaches for bad posture detection using computer vision and IoT, *E3S Web of Conferences*, 507, 01045. https://doi.org/10.1051/e3sconf/202450701045

[15] Stefana, E., Marciano, F., Rossi, D., Cocca, P., & Tomasoni, G. (2021). Wearable Devices for Ergonomics: A Systematic Literature Review. *Sensors*, 21(3), 777. https://doi.org/10.3390/s21030777

[16] Yolo. (n.d.). sleep1 Object Detection Dataset and Pre-Trained Model. *Roboflow*. https://universe.roboflow.com/yolo-i9xsa/sleep1

[17] Yu, J., & Zhang, W. (2021). Face Mask Wearing Detection Algorithm Based on Improved YOLO-v4. *Sensors*, 21(9), 3263. https://doi.org/10.3390/s21093263

[18] Zhang, Y., Guo, Z., Wu, J., Tian, Y., Tang, H., & Guo, X. (2022). Real-Time Vehicle Detection Based on Improved YOLO v5. *Sustainability*, 14(19), 12274. https://doi.org/10.3390/su141912274