ISSN: 2581-1274

Implementation of YOLO (You Only Look Once) Algorithm for Drowsiness Detection as An Additional Safety Feature in the Operation of Crane Equipment in Real Time

Riza Adrian Maulana [a,1,*], Hardiansyah [a,2]

[a]*University of Pamulang, Jl. Raya Puspitek, Kec. Pamulang, Kota Tangerang Selatan, Banten 15310, Indonesia*
[1] *Rizaaaawww28@gmail.com**; [2] *dosen02058@unpam.ac.id*

ARTICLE INFO

Article history:
Published

Keywords:
Detection System
Yolo Algorithm
Drowsiness
Construction Workers

ABSTRACT

Work accidents, especially in the construction sector, are still a serious problem, with fatigue as one of the main causes. Based on data from Kemnaker, the author realizes the need for early prevention solutions to reduce the risk of accidents due to fatigue. One of the approaches proposed is the development of an automatic detection system to recognize workers' facial expressions, especially in detecting levels of freshness and sleepiness. The obstacles that are often faced are limited time and scale in manual monitoring, especially on large-scale construction projects. To overcome this, the You Only Look Once (YOLO) algorithm is used, which is able to detect objects quickly and accurately, to provide continuous monitoring of workers' conditions. This research focuses on the application of the YOLOv8n model in an automatic freshness and sleepiness facial expression detection system. The model is trained using a dataset that includes a variety of facial expressions in different situations, allowing the system to detect worker conditions in real-time and at scale. The evaluation results in this research show very good performance, with precision reaching 99.9%, recall 100%, mAP50 99.5%, and mAP50-90 97.9%. Although the model sometimes makes mistakes in object class recognition, the overall results still show a very high level of accuracy. With this system, it is hoped that it can improve work safety through early detection of signs of fatigue in workers, so that the potential for work accidents can be significantly minimized.

I. Introduction

Large equipment such as cranes require a very high level of concentration and caution from the operator. Every small mistake in the company can be fatal not only for the operator himself, but also for other employees and the entire work environment [1]. Accidents involving cranes often result in considerable material losses, such as: Damage to infrastructure and equipment, as well as the potential for countless deaths. In this case, one of the main factors that cause accidents is the condition of the drowsy operator [2]. When operators feel sleepy, their alertness and reaction time decrease, which can lead to poor decisions. Therefore, it is very important to recognize and fix this sleepiness problem to improve safety in the workplace. Crane leaders often work long and monotonous, significantly increasing the risk of fatigue and drowsiness. An unsupportive work environment, such as a lack of breaks or the right pressure to perform a task in a short period of time, can exacerbate this situation [3]. Traditional methods of monitoring operator status, such as direct observation by superiors, have proven ineffective. Manual monitoring cannot issue actual alerts. It is often based on subjective assessments and can lead to delayed responses to signs of fatigue. Therefore, an automated system is needed to quickly and quickly recognize a drowsy operator to alert you early and prevent accidents [4].

DOI: 10.31572/inotera.Vol10.Iss1.2025.ID458

This study shows that you apply the implementation of the algorithm (learning) only once to recognize the drowsiness of the crane leader in real time. Yolo is a deep learning algorithm that reliably and efficiently recognizes objects, including facial features that indicate drowsiness, such as closed eyes, blinking, and head position [5]. By training the algorithm using data records from facial images marked "sleepy" and "sleepy", the system can analyze the video input from the camera pointed at the crane guide surface. By detecting drowsiness early, the system can be a warning for the operator or even an automatic stop of crane operations to avoid accidents. Therefore, the goal of this study is to develop additional safety features that improve work safety in the crane company environment and provide better protection for operators and other employees

II. Method

A. *Data Collection Method*

In this study, the researcher used a quantitative research method. According to [6] the quantitative research method refers to a research approach based on the philosophy of positivism. This method is used to investigate a specific population or sample, through data collection using research instruments and data analysis that is quantitative or statistical, with the aim of testing hypotheses that have been formulated. In this study, the method applied is quantitative because the development of the YOLO method requires the use of a large dataset for training and testing purposes [7]. This approach allows for large-scale data analysis by taking into account large amounts of data, as well as facilitating decision-making based on empirical data and statistical analysis.

B. *You Only Look Once (YOLO)*

The YOLO method means that the input image is divided into smaller networks. This makes each grill responsible for recognizing objects within that area. Each grid predicts many boxes and bounding probabilities for all the objects in it. This process includes several important steps [8][9]:

1. Image Sharing
   The input image is divided into a fixed-sized grid, e.g. 7x7. Each cell in this grid will predict the objects that are within that area.

2. Bounding Box Prediction
   Each grid cell predicts multiple bounding boxes that include objects, as well as a confidence score that indicates how confident the model is that the object is in that box. This confidence value is calculated based on the accuracy of the prediction and the size of the object.

3. Object Classification
   In addition to predicting the bounding box, each cell also predicts the class of objects that may be in the box. This is done by using the SoftMax activation function to generate class probabilities.

4. Non-Maximum Suppression
   After all the predictions are made, YOLO applies a non-maximum suppression technique to eliminate overlapping bounding boxes and keep only the predictions with the highest confidence value. This helps to reduce duplication and improve detection accuracy.

III. Results and Discussion

This process includes various stages from data collection and labelling, pre-processing, to model training using the YOLOv8n algorithm. This explanation will provide an in-depth overview of the approaches and methodologies that will be used in this study. As such, readers will gain a comprehensive understanding of how this model was developed and tested to effectively detect sleepy and fresh expressions.

A. Dataset Collection

The dataset used in this study was made by the researcher by taking the existing dataset in the kaggle.com to be used as a dataset in this study. The collection of this dataset aims to train the YOLO algorithm in detecting sleepy and fresh facial expressions. The total images collected were 1000 images, consisting of 500 images of sleepy faces and 500 images of fresh faces. With this balanced dataset, it is hoped that the YOLO algorithm can learn and recognize the difference between sleepy faces and fresh faces with a high level of accuracy.

Fig 1. Sleepy and Fresh Dataset

The figure above shows some examples of datasets consisting of sleepy and fresh faces. Each image in this dataset shows a variety of facial expressions, ranging from very fresh to those that appear distinctly sleepy. These examples reflect different levels of facial expressions, which aims to ensure the model can recognize accurate differences. By looking at these image examples, it can be understood how the dataset is designed to train the YOLO algorithm in detecting sleepy and fresh facial expressions effectively.

B. Dataset labeling

Headings Here is a schematic from the labeling of the data until the system is able to detect the object.

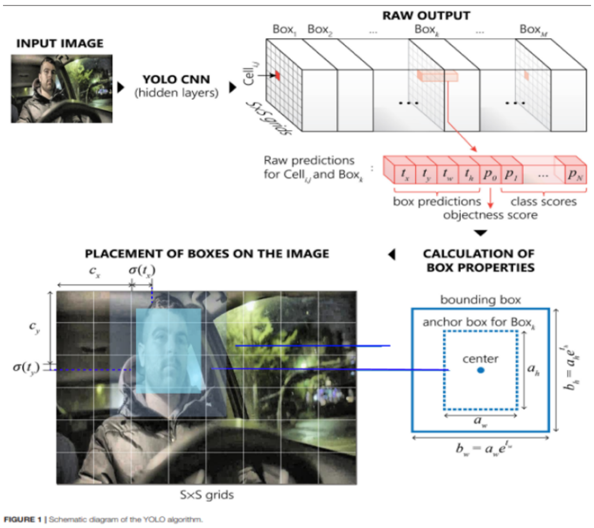

Fig 2. YOLO Schematic Diagram

The image above uses the Roboflow platform in its labelling process. Each image in the dataset, which consists of sleepy and fresh faces, is uploaded to Roboflow for further processing. This labelling process involves marking important areas on the image that indicate the condition of the facial expression. Using the tools available in Roboflow, researchers can easily tag and classify each image according to a predefined category.

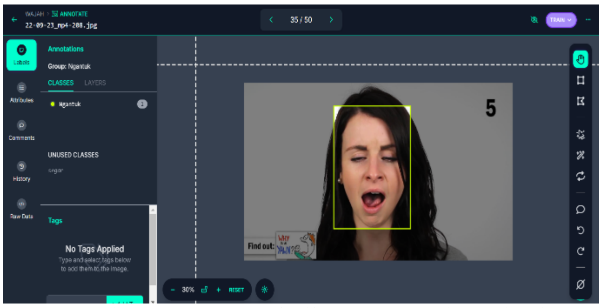

Fig 3. Dataset Labelling Process in Roboflow

C. Dataset Sharing

Once the labeling process is complete using Roboflow for the facial expression dataset, the next step is to divide the dataset into sections for model training, validation, and testing. This dataset division is important to objectively test the performance of the model and ensure good generalization in detecting facial expressions. The initial dataset consisted of 1000 images of sleepy and fresh faces and then separated using Roboflow with a ratio of 85% for training data, 10% for validation data, and 5% for test data. This process allows researchers to train the YOLOv8n model with Most of the data.

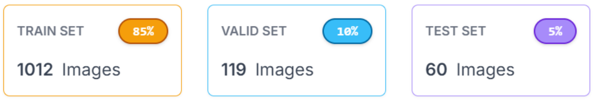

Fig 4. Split Dataset

The image above has an initial dataset consisting of 1000 images of facial expressions that have undergone the augmentation process, so the total number of images in the dataset increases to 1191 images. This augmentation is carried out to enrich the variety of data used in model training.

D. Dataset Augmentation

The Augmentation process is carried out on a dataset of fresh facial expressions to enrich the variety of available data. This augmentation includes several methods such as flip, crop, rotation, etc. By using some additional variation techniques, it allows the model to get to know objects with various conditions better.

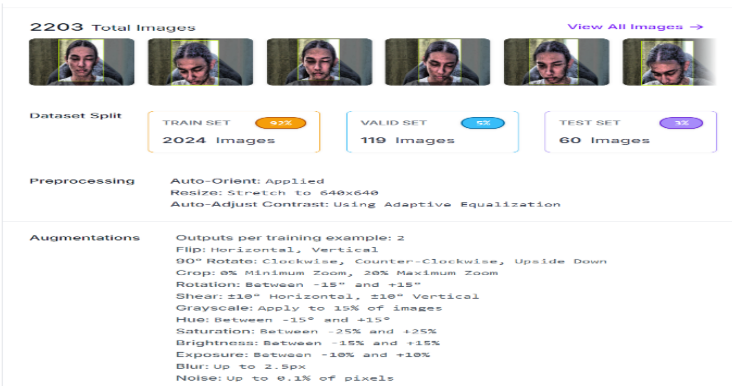

Fig 5. Dataset Augmentation Results in Roboflow

With the application of these various augmentation methods, the dataset that initially consisted of 1000 images increased to 2024 images. The augmentation process causes the images to multiply, resulting in a richer and more diverse dataset. This aims to ensure that the YOLOv8n model can be trained more effectively, being able to recognize and classify sleepy and fresh facial expressions with higher accuracy.

E. Test Model YOLO

The YOLO Model Test is the result of dataset testing conducted using the YOLOv8n model. In this test, the dataset consisted of 60 samples, 30 fresh face samples and 30 sleepy face samples. The model was tested to detect the entire test data, and the results showed that the YOLOv8n model could perform the detection well. Accurate detection of these two types of qualities confirms the effectiveness and ability of the model to identify test data objects well. The YOLO model used in this study successfully predicted the images and videos taken live accurately, showing a very high level of confidence in each class. However, it should be noted that not all images or videos can be predicted well by this model, especially images or videos taken from Google. This is due to the limitations of the dataset used in this study, which may not cover a wide enough variety of images or videos to cover all possible conditions in the field.

Fig 6. Test Model YOLO

F. Accuracy of Result

In this stage, we will discuss the analysis of the accuracy of the results obtained from the YOLOv8n model training matrix. This analysis includes the evaluation of performance metrics such as precision, recall, and F1-score to measure how well the model can detect and classify fresh and sleepy facial expressions. By examining the values of this matrix, we can understand the strengths and weaknesses of the model, as well as determine whether it meets the desired criteria for practical applications. A detailed explanation of the performance metrics and interpretation of these results will provide a thorough picture of the model's effectiveness in facial expression detection tasks.

1. F1-Score

   The F1-Confidence *curve* is used to evaluate the performance of the classification model. On the horizontal axis is the *confidence* value, and on the vertical axis is *the F1-score* value. This curve illustrates the relationship between the prediction confidence level and the model's performance. The two classes evaluated in this curve are "sleepy" (depicted in light blue) and "fresh" (depicted in orange). In addition, the curve for "*all* classes" is shown with a bold blue line with an *F1-score* of 1.00 at a *confidence* level of 0.5. This indicates that the model is performing well with a high *F1-score* close to 1 at that *confidence* level .

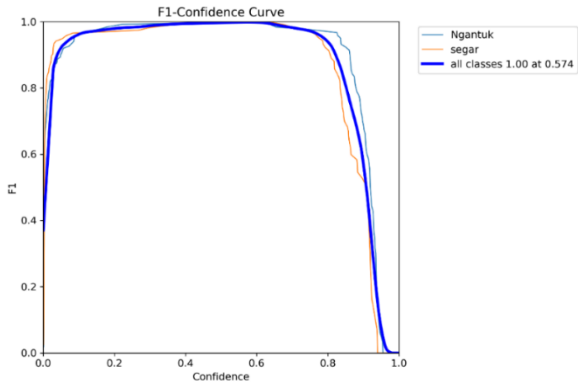

Fig 7. F1-Score

2. Precision
   Precision-Confidence *curve* that evaluates the performance of the classification model. This curve shows the relationship between the prediction confidence level (horizontal axis) and *precision* (vertical axis) for the "sleepy" (light blue) and "fresh" (orange) classes. The "all classes" curve (thick blue line) reaches *a precision* of 1.00 at 0.660 confidence, indicating good performance. *Precision* is close to 1 for most *confidences*, indicating good classification ability.

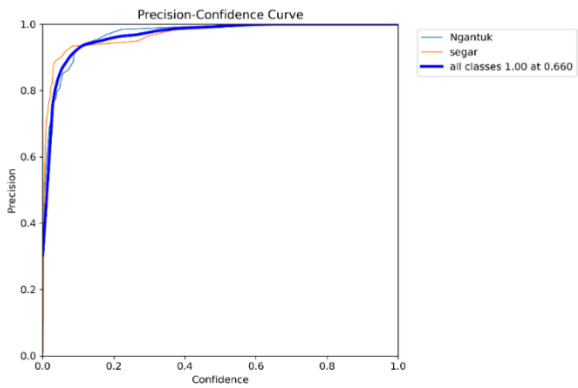

Fig 8. Precision

3. Precision Recall
   The precision and recall values for the prediction have good results, with the best precision and recall values around 0.995. This means that the model is able to predict many objects with a high degree of accuracy and precise labels. This model shows near-perfect performance in identifying and classifying objects according to the correct labels.

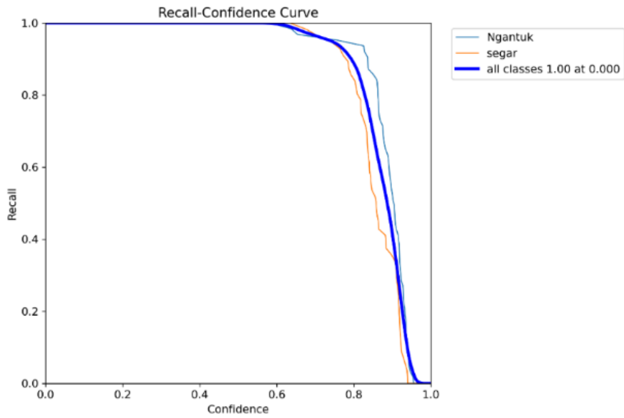

Fig 9. Precision Recall

4. Confusion Matrix
   The confusion matrix features two classes, namely "rotten" and "fresh". In this matrix, the horizontal axis (True) indicates the actual class, while the vertical axis (Predicted) indicates the class predicted by the model. The values on each matrix cell indicate the correct prediction of the training results. The model has perfect classification performance for each class. This is indicated by a 100% value on each class, which means that all the model predictions for each class are correct. No classification errors occurred because the values outside the diagonal were all 0. The model gives quite satisfactory results in each class. Thus, the model can be relied upon to provide consistent and accurate results in the classification of the two classes.

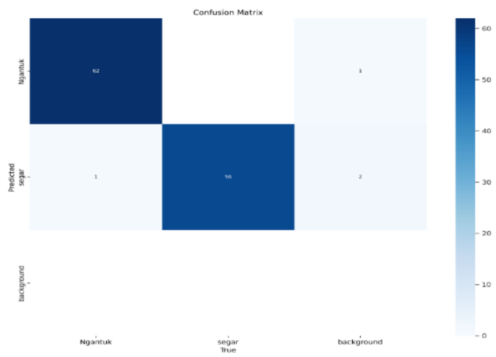

Fig 10. Confusion Matrix

IV. Conclusion

The application of the YOLO (You Only Look Once) fatigue detection algorithm as an additional safety feature in real-time crane operations shows the potential to significantly improve occupational safety in high-risk operating environments. YOLO's ability to detect objects quickly and accurately allows the system to provide early warning to the crane operator if signs of fatigue or drowsiness, such as closed eyes or reduced flickering, are detected. The results of the evaluation in this study showed excellent performance, with precision reaching 99.9%, recall 100%, mAP50 99.5%, and mAP50-90 97.9%. Although the model sometimes makes mistakes in object class recognition, the overall results still show a very high level of accuracy. With this system, it is hoped that it can improve work safety through early detection of signs of fatigue in workers, so that the potential for work accidents can be significantly minimized.

Vol. 10, No. 1, January-June 2025, pp. 120-120

References

[1] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2016, pp. 779–788. doi: 10.1109/CVPR.2016.91.

[2] R. Bhardwaj and H. Singh, "Modeling the Effects of Intellectual Capital on Decision Making - A Study of Interaction Moderation with Knowledge Management Process," *Int. J. Comput. Appl.*, vol. 180, no. 35, pp. 37–50, 2018, doi: 10.5120/ijca2018916885.

[3] M. Hadavi, Z. Ghomian, F. Mohammadi, and A. sahebi, "Workplace violence against health care workers during the COVID-19 Pandemic: A systematic review and meta-analysis," *J. Safety Res.*, vol. 85, pp. 1–7, 2023, doi: https://doi.org/10.1016/j.jsr.2023.01.001.

[4] M. Sarosa and N. Muna, "Implementasi Algoritma You Only Look Once (Yolo) Untuk Implementation of You Only Look Once (Yolo) Algorithm for," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 8, no. 4, pp. 787–792, 2021, doi: 10.25126/jtiik.202184407.

[5] I. M. D. Maleh, R. Teguh, A. S. Sahay, S. Okta, and M. P. Pratama, "Implementasi Algoritma You Only Look Once (YOLO) Untuk Object Detection Sarang Orang Utan Di Taman Nasional Sebangau," *J. Inform.*, vol. 10, no. 1, pp. 19–27, 2023, doi: 10.31294/inf.v10i1.13922.

[6] K. S. S, I. Taufik, I. Ramadhani, and A. A. Siregar, "Implementation of You Only Look Once Version 8 Algorithm to Detect Multi-Face Drivers and Vehicle Plates," vol. 11, no. 2, pp. 103–108, 2024.

[7] Q. Aini, N. Lutfiani, H. Kusumah, and M. S. Zahran, "Deteksi dan Pengenalan Objek Dengan Model Machine Learning: Model Yolo," *CESS (Journal Comput. Eng. Syst. Sci.*, vol. 6, no. 2, p. 192, 2021, doi: 10.24114/cess.v6i2.25840.

[8] A. Wibowo, L. Lusiana, and T. K. Dewi, "Implementasi Algoritma Deep Learning You Only Look Once (YOLOv5) Untuk Deteksi Buah Segar Dan Busuk," *Paspalum J. Ilm. Pertan.*, vol. 11, no. 1, p. 123, 2023, doi: 10.35138/paspalum.v11i1.489.

[9] L. Rahma, H. Syaputra, A. H. Mirza, and S. D. Purnamasari, "Objek Deteksi Makanan Khas Palembang Menggunakan Algoritma YOLO (You Only Look Once)," *J. Nas. Ilmu Komput.*, vol. 2, no. 3, pp. 213–232, 2021, doi: 10.47747/jurnalnik.v2i3.534.

10. Rater Salya,Balkhaya "The Influence of Social Media Digital Marketing Analysis on New Student Acceptance in a University" J.Inotera,vol.8, no. 1 PP. 150-154, 2023, doi: 10.31572/inotera.Vol8.Iss1.2023.ID226