JITK (JURNAL ILMU PENGETAHUAN DAN TEKNOLOGI KOMPUTER)

VOL. 10. NO. 1 AUGUST 2024
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v10i1.5576

PERFORMANCE OF THE YOLOV5 ALGORITHM TO DETECT HUMANS IN THE REAR EXCAVATOR AREA

Hanna Naili Syifa'[1]; Anan Nugroho[2*]

Electrical Engineering[1,2]
Universitas Negeri Semarang, Indonesia[1,2]
https://unnes.ac.id/[1,2]
hannasyifa06@students.unnes.ac.id[1], anannugroho@mail.unnes.ac.id[2*]

(*) Corresponding Author
(Responsible for the Quality of Paper Content)

Abstract*—Work involving excavators carries a high risk of accidents that can result in fatalities, making Occupational Safety and Health (OSH) critically important. Most excavator accidents are caused by blind spots at the rear, where the operator's limited field of view increases the risk of hitting nearby objects or workers. Despite safety features such as reverse alarms and rear cameras, these technologies only display real-time video without automatically detecting workers, thus still posing a significant risk. This study aims to develop a human detection system for the rear area of excavators using the YOLOv5 algorithm based on image processing. The system's main features include real-time human detection, distance estimation, and audible warnings if a human is detected within a high-risk distance. The system was tested using three video recordings depicting human objects behind the excavator in different scenarios. Despite the limited number of video samples, the human objects provided sufficient complexity to evaluate the system's effectiveness. The test results showed an average accuracy of 80.5% and an F1-score of 87.78%. These findings indicate that the YOLOv5-based detection system performs well in various video conditions and shows potential effectiveness in real operational situations. Consequently, this system is expected to reduce the risk of work accidents with excavators caused by rear blind spots and improve on-site worker safety. This research contributes to the field of occupational safety by integrating image processing algorithms into the development of heavy equipment safety technology, thereby enhancing worker protection.*

Keywords*: blind spot, excavator, human detection system, worker safety, YOLOv5 algorithm.*

Intisari—Pekerjaan yang melibatkan ekskavator memiliki risiko tinggi terhadap kecelakaan yang dapat berakibat fatal, sehingga Keselamatan dan Kesehatan Kerja (K3) sangat penting. Sebagian besar kecelakaan eskavator disebabkan oleh titik buta di bagian belakang, di mana pandangan terbatas operator meningkatkan risiko menabrak objek atau pekerja di sekitar. Meskipun fitur keselamatan seperti alarm mundur dan kamera belakang sudah ada, teknologi ini hanya menampilkan video secara real-time tanpa kemampuan mendeteksi keberadaan pekerja secara otomatis, sehingga masih menimbulkan risiko signifikan. Penelitian ini bertujuan untuk mengembangkan sistem deteksi manusia di area belakang ekskavator menggunakan algoritma YOLOv5 berbasis pengolahan citra. Fitur utama sistem ini mencakup deteksi manusia secara real-time, estimasi jarak, dan peringatan suara jika manusia terdeteksi dalam jarak berisiko tinggi. Sistem ini diuji menggunakan tiga rekaman video yang memperlihatkan objek manusia di belakang ekskavator dalam skenario yang berbeda. Meskipun jumlah sampel video terbatas, objek manusia memberikan kompleksitas yang cukup untuk mengevaluasi efektivitas sistem. Hasil pengujian menunjukkan akurasi rata-rata sebesar 80,5% dan skor F1 sebesar 87,78%. Temuan ini menunjukkan bahwa sistem deteksi berbasis YOLOv5 memiliki kinerja yang baik dalam berbagai kondisi video dan menunjukkan potensi efektivitas dalam situasi operasional nyata. Dengan demikian, sistem ini diharapkan dapat mengurangi risiko kecelakaan kerja dengan ekskavator yang disebabkan oleh titik buta di bagian belakang dan meningkatkan keselamatan pekerja di lokasi. Penelitian ini berkontribusi pada bidang keselamatan kerja dengan mengintegrasikan algoritma pengolahan citra ke dalam pengembangan teknologi keselamatan alat berat, sehingga meningkatkan perlindungan pekerja.

Kata Kunci: *titik buta, ekskavator, sistem deteksi manusia, keselamatan pekerja, algoritma YOLOv5.*

INTRODUCTION

Work involving excavators carries a high risk of workplace accidents that can result in fatalities, making Occupational Safety and Health (OSH) of workers critically important. This is regulated in the Minister of Manpower Regulation of the Republic of Indonesia No. 8 of 2020 concerning Occupational Safety and Health for Lifting and Transport Equipment in Chapter 1, Article 1, Paragraph 1, which states that OSH encompasses all activities to ensure and protect the safety and health of workers through the prevention of workplace accidents and occupational diseases [1]. Excavator-related accidents have caused several fatal incidents for workers. In 2021, a worker was crushed by an excavator's track at PT Indonesia Weda Bay Industrial Park (IWIP) because the operator was unaware of the victim's presence behind the machine [2]. Similar incidents occurred at Pasar Induk Cibitung, Bekasi, in 2023 [3], and at an illegal mining site in Sukokerto Village, Jember, East Java, in 2023 [4]. Most excavator accidents are caused by blind spots at the rear, which are areas that cannot be seen by the operator either directly or through mirrors [5]. These blind spots occur due to the large size of the excavator, limiting the operator's visibility and increasing the risk of hitting objects or workers nearby [6].

Worker safety is a top priority in OSH because most workplace accidents around excavators involve workers [7]. Although excavators are equipped with built-in safety features such as reverse alarms and rear cameras, these technologies have limitations as they only display real-time video on the monitor screen without automatic detection of workers' presence [8]. This limitation increases the risk of accidents, especially due to the large blind spots on excavators, which cause operators to be unaware of workers nearby. This risk can result in serious injuries or fatalities. In this context, advances in image processing technology based on deep learning can be a potential solution to develop safety features on excavators with the ability to automatically detect human objects and determine their distance [9]. This technology can enhance awareness of blind spots and reduce accident risks. Compared to other sensor technologies such as Radio Frequency Identification Technology (RFID), The Global Positioning System (GPS), and Ultra-Wideband (UWB), deep learning-based image processing technology does not require the installation of additional sensors on vehicle components, making it more cost- and time-efficient [10]. Additionally, the quick and accurate detection capabilities of this technology have great potential to be implemented as a safety feature on excavators [11]. Therefore, this research aims to develop a human detection system in the rear area of the excavator using deep learning-based image processing technology to improve workplace safety.

Previous research on human object detection for safety management at construction sites has employed various approaches. The use of Mask Region-based Convolutional Neural Network (Mask R-CNN) algorithms and gradient-based methods has effectively improved worker detection [12]. Faster Region-based Convolutional Neural Network (Faster R-CNN) algorithms and Long Short-Term Memory (LTSM) network have been employed for real-time worker detection, facilitating the monitoring of unsafe conditions and behaviors at construction sites [13]. Another approach integrates You Only Look One version 4 (YOLOv4) and Siamese network algorithms to detect and track workers in real-time, significantly reducing false alarms and enhancing safety in heavy equipment operations [14]. These approaches demonstrate that deep learning-based object detection methods can improve workplace safety more accurately, efficiently, and specifically at construction sites. Deep learning-based object detection is divided into two categories: single-stage detection, such as YOLO, and two-stage detection, such as R-CNN, Faster R-CNN, and Mask R-CNN [15]. Although single-stage detection shows lower accuracy compared to two-stage detection, it performs better in terms of detection speed, making it a more efficient choice for real-time applications [16]. YOLOv3 has better object detection capabilities compared to previous versions, but it often suffers from high background errors, especially in complex construction sites [17]. YOLOv5 addresses this issue through mosaic data augmentation during the training process, enabling effective learning even with limited data [15]. With smaller training image sizes, YOLOv5 is better at detecting small objects and is suitable for large-scale construction site monitoring due to its high accuracy and fast detection speed compared to other YOLO series [18]. Recent studies that integrate YOLOv5

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

VOL. 10. NO. 1 AUGUST 2024
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v10i1.5576

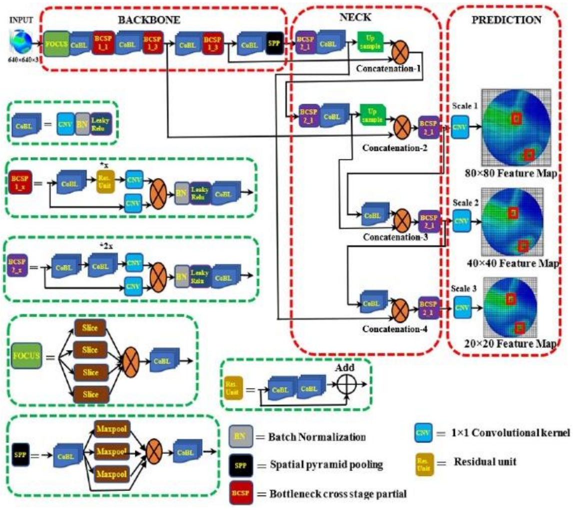
algorithms for real-time detection of workers and heavy equipment have enhanced comprehensive construction site monitoring and safety management [18], [19]. Based on these findings, YOLOv5 has the potential to improve human detection efficiency for safety features in the rear area of the excavator.

In this research, the YOLOv5 algorithm will be used to detect the presence of humans in the blind spot area behind the excavator. YOLOv5 is a neural network-based object detection algorithm that belongs to the single-stage detection network [20]. This algorithm was chosen for its high detection speed and accuracy, making it suitable for real-time object detection applications [21]. YOLOv5 can detect objects at speeds of up to 140 FPS and has a smaller file size compared to previous versions [22]. Additionally, YOLOv5 has efficient reasoning capabilities for various types of direct input such as single images, image batches, videos, and webcam input [23]. The advantages of YOLOv5 in quickly and accurately detecting objects make it an ideal choice for improving workplace safety on excavators by minimizing the risk of accidents caused by blind spots.

This research is expected to make a significant contribution to the development of safety technology on excavators. By integrating a YOLOv5-based human detection system into excavator operations, the system will recognize and track the presence of humans in the rear area of the excavator by providing detection information such as bounding boxes, class labels, confidence scores, and distance information displayed on the operator's monitor screen. Additionally, the system will provide real-time audio warnings to the operator when a human object is detected within a potentially dangerous distance. This warning helps the operator take preventive actions to avoid workplace accidents, thereby enhancing safety in the excavator's operational area.

MATERIALS AND METHODS

This research aims to detect human objects and calculate their distance in the rear area of excavators using the YOLOv5 algorithm, which is one of the popular algorithms in neural network-based object detection [24]. YOLOv5 was developed with the goal of improving speed and accuracy in object detection [25], and it utilizes an end-to-end neural network to directly generate bounding boxes and class labels [26]. YOLOv5 consists of three core components: the backbone, neck, and prediction head [27]. The backbone employs Bottleneck Cross Stage Partial Darknet (BCSPD) to extract initial visual features from images, while the neck utilizes Path Aggregation Network (PANet) to merge information from various resolution levels to enhance detection accuracy. The prediction head produces three different features to perform multi-scale detection on small, medium, and large objects, including bounding boxes. The architecture of YOLOv5 is depicted in Figure 1.



Source: (YOLOv5 Architecture [27], 2022)
Figure 1. YOLOv5 Architecture

Based on Figure 1, there are several stages in YOLOv5 for object detection [28]. The first stage begins by applying a 640x640-sized image into the backbone. Then, the image is divided into four small parts by the FOCUS module. Through a series of convolution operations and BCSP1, the feature map is moved to the second merging layer. On the other hand, after performing one BCSP1 execution, two BCSP2 executions, convolution operations, and two upsampling processes, a feature map of size 80x80 is obtained. In the second stage, the 80x80 feature map is processed with a 3x3 convolution kernel and directed towards the third merging layer. The previous upsampling feature map is run by a single 1x1 convolution kernel and directed towards the third merging layer. After completing the BCSP2 layer and a single 1x1 convolution operation, a feature map of size 40x40 is obtained. In the third stage, the 40x40 feature map is processed by a 3x3 convolution kernel and directed towards the fourth merging layer. Additionally, the previous upsampling feature map is run by a single 1x1 convolution kernel and directed towards the fourth merging layer. After performing BCSP2 and 1x1 convolution operations, a feature map of size 20x20 is obtained. In the object identification stage, feature maps of different sizes are responsible for identifying objects. The model predicts the location of the target object with bounding boxes and confidence scores.

In object detection using YOLOv5, there is a possibility that multiple bounding boxes detect the same object. To address this, the Non-Maximum Suppression (NMS) method is used. NMS functions to reduce duplications in detection results by removing overlapping bounding boxes based on the confidence score percentage [28]. In its implementation, YOLOv5 first selects the bounding box with the highest confidence score. Next, YOLOv5 compares this bounding box with all other bounding boxes from the same class using Intersection over Union (IoU). If the IoU between two bounding boxes exceeds a predetermined threshold value, such as 0.5, then the bounding box with the lower confidence score is removed [29]. This process is repeated until all bounding boxes have been evaluated. Next, the bounding boxes are used as crucial parameters in the process of calculating the distance between the camera and the detected object.

In the process of distance calculation using images, information from the bounding boxes such as width (w), height (h), and coordinates of the midpoint (x, y) of the detected object are used as inputs to calculate the distance of the object from the camera using equation (1) [30].

$$d = \left(\frac{2x3.14x180}{w+hx360}\right)x1000 + 3 \qquad (1)$$

Based on equation (1), the calculation process involves converting the camera's field of view into the focal length, often referred to as the arc length on the circle. By leveraging the camera's field of view and information from the bounding box, the actual distance of the object from the camera can be calculated [31]. For example, the larger the size of the object in the image, the closer the object's distance. Conversely, the smaller the size of the object in the image, the farther the object's distance from the camera.

The hardware and software used in this research are as follows:

1. Lenovo IdeaPad 330 laptop with the following specifications:
   a. Processor: Intel® Celeron® N4100 CPU @1.10GHz
   b. Graphics: Intel® UHD Graphics 600
   c. RAM: 4 GB
   d. Storage: 128 GB SSD
   e. OS: Windows 10 Home
2. External Webcam with 1080p (Full High Definition) resolution
3. 15-meter long USB 2.0 Male to Female cable
4. Software used includes:
   a. Visual Studio Code (Text editor)
   b. Google Colaboratory (Cloud computing-based notebook)
   c. Roboflow (Dataset annotator)
   d. Python 3.11.9 Programming Language
   e. OpenCV 4.9 (Library)

In this research, the materials used are three video recordings of human objects behind an excavator with a duration of 10 seconds each in different scenarios. Video scenarios 1 and 2 show several stationary human objects with varying positions and distances. Meanwhile, scenario video 3 shows several moving human objects with varying positions and distances. Although the dataset used is relatively limited, the human objects in the video have enough complexity variations to test the system's effectiveness and make the research focus on in-depth analysis of specific scenarios faced in the excavator working environment. Using this dataset provides essential initial understanding in developing accurate detection and distance calculation methods for field situations. However, it is highly recommended to expand the dataset with more and diverse videos to improve the model's ability to adapt to various real-life situations.

The stages of this research are as follows:

Dataset Acquisition

In this research, the dataset consists of human images with various orientations annotated from Roboflow. The dataset includes 2,142 images for use in training the YOLO model. This dataset is divided into three types of data: 1,502 images for training data, 426 images for validation data, and 214 images for testing data. The annotation process involves manually labeling human objects in each image to ensure accurate detection. Roboflow provides tools to facilitate this labeling process, ensuring that each human object is correctly annotated with bounding boxes. This annotation is crucial for informing the model of the exact location of human objects in the images.

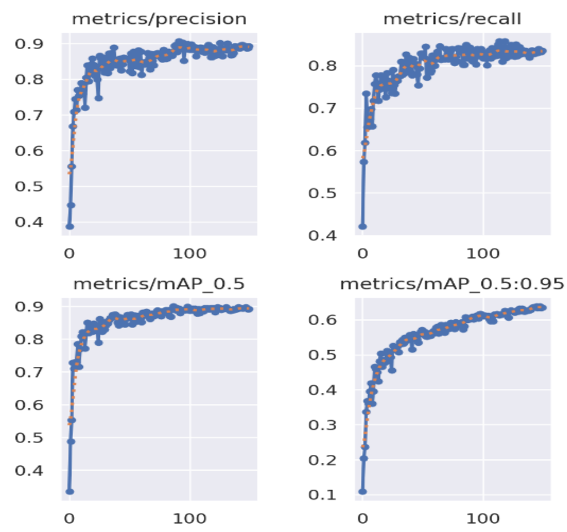
Training Data

In this research, training data is conducted to generate model weights used for human detection. The training data process uses the YOLOv5 algorithm with PyTorch on Google Colaboratory. Training lasted for one hour with several specified parameters: the image size used is 640x640 pixels to provide better detail in human object detection, and the batch size used is 32 to increase training speed by processing more images simultaneously. The training is conducted for 150 epochs to allow the model to learn more complex patterns in the

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

VOL. 10. NO. 1 AUGUST 2024
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v10i1.5576

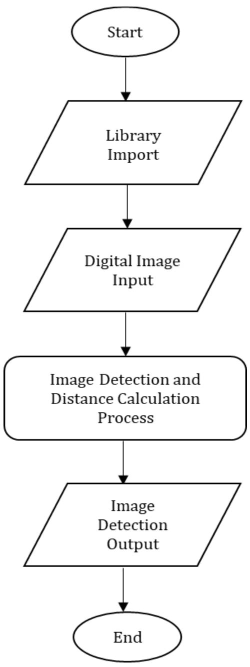
data. The model is trained to detect one class of objects, namely humans. After the training data is completed, the YOLOv5 model generates a file.onnx file to be implemented in the system for human detection. The YOLOv5 training results show a precision value of 0.973, recall of 0.93, F1-score of 0.86, and mAP of 89.3%. These results indicate good performance of the model in detecting humans, as show in Figure 2.



Source: (Research Results, 2024)
Figure 2. YOLOv5 Training Results

System Implementation

In this research, the detailed workflow of the human detection system can be seen in Figure 3.



Source: (Research Results, 2024)
Figure 3. Workflow of Human Detection System

Figure 3 illustrates several steps employed by the human detection system in the rear area of an excavator based on video camera, including:

1. Library Import Stage
   At this stage, the Open-Source Computer Vision (OpenCV) library is imported for processing images, videos, and real-time information extraction. OpenCV can recognize and detect objects in complex situations [32].
2. Digital Image Input Stage
   At this stage, digital images in the form of videos are input. The input video consists of three 10-second videos, each with different scenarios in (.mp4) file format.
3. Image Detection and Distance Calculation Stage
   At this stage, human objects are detected in the input video using the YOLOv5 algorithm, and the distance from each detected human object is calculated. The process begins by applying 640x640 sized images to the backbone and dividing them into four small parts by the FOCUS module. Through a series of convolution operations and BCSP, feature maps are generated at various scales (80x80, 40x40, and 20x20). The model then predicts the object locations with bounding boxes and confidence scores. To address duplicate detections, the NMS method removes overlapping bounding boxes based on confidence scores. After obtaining the correct bounding boxes, the distance between the detected human objects and the excavator is calculated using Equation (1), where the width (w) and height (h) of the bounding box in pixels are used for accurate distance calculation. This process involves converting the camera's Field of View into the arc length on a circle. By utilizing the camera's Field of View and information from the bounding box, the actual distance of human objects in the rear area of the excavator is determined.
4. Image Detection Output Stage
   At this stage, the output of the detection process includes bounding boxes, class labels, confidence scores, and distance information. Human objects detected at a distance of 3 meters or more are marked with green bounding boxes, while human objects detected at a distance of less than 3 meters are flagged as high-risk situations with red bounding boxes and accompanied by a warning sound. High-risk situations indicate that human objects are detected in close proximity to the excavator, posing potential safety hazards.

System Testing

In this research, testing is conducted to measure the system's accuracy in detecting humans and their distances in real-time. The testing process involves evaluating how the detection results in input videos from an external webcam are divided into three groups: True Positive (TP), False Positive (FP), and False Negative (FN) [33]. TP refers to bounding boxes in the positive class that result in correct detections. In this case, human objects detected at a distance of 3 meters or more are displayed with green bounding boxes, while human objects detected at a distance of less than 3 meters are shown as high-risk with red bounding boxes. FP refers to bounding boxes in the positive class that result in incorrect detections. In this case, objects are detected but incorrectly, such as when no human is present but detected as human. Additionally, if human objects detected at a distance of 3 meters or more are displayed as high-risk with red bounding boxes, while human objects detected at a distance of less than 3 meters are displayed with green bounding boxes. FN refers to bounding boxes in the negative class that result in incorrect detections. In this case, if there are humans but not detected as humans. Based on the TP, FP, and FN values obtained, accuracy and F1-score are calculated to determine the detection performance according to Equations (2) and (3) [34].

$$Accuracy = \frac{TP}{TP+FP+FN} x100\% \tag{2}$$

$$F1 - score = \frac{2xPrecisionxRecall}{Precision+Recall} \tag{3}$$

The selection of accuracy and F1-score as performance indices for detection is based on several reasons. Accuracy and F1-score are the most popular metrics for evaluating model performance across various machine learning tasks [35]. Accuracy indicates the percentage of correct predictions out of all predictions made and provides a general indication of model performance. This method is often used due to its ease of interpretation [36]. On the other hand, F1-score provides a more detailed assessment of model performance by combining precision and recall, which is particularly fair under class imbalance conditions [37]. Precision measures the proportion of true positive detections out of all positive predictions, crucial in occupational safety contexts to minimize false positive detections [36]. Meanwhile, recall measures the model's ability to detect all existing positive cases, essential to ensure all humans present are detected, thereby reducing accident risks [35]. By combining accuracy and F1-score, a comprehensive and detailed picture of human detection performance is obtained, ensuring that the model is not only accurate overall but also effective in correctly detecting humans and avoiding risky detection errors in workplace safety behind excavators.

RESULTS AND DISCUSSION

Image Input

Observations in this study were conducted by observing detection results on a digital image input in the form of a video. The video used was captured by an external 1080p resolution webcam (Full High Definition) mounted behind an excavator. The external webcam recorded videos at 60 frames per second. The videos depicted humans in various conditions behind the excavator, with each video lasting 10 seconds, resulting in 600 frames per video. These videos were input into the developed program in (.mp4) format. An illustration of video capture can be seen in Figure 4.

Source: (Research Results, 2024)

Figure 4. Illustration of Video Capture

Detection Results

In this research, video detection was solely focused on human objects, thus other objects present in the videos were not detected. Detection was performed on three recorded videos of humans in the area behind the excavator with different scenarios, each consisting of 600 frames. Video scenarios 1 and 2 show several stationary human objects with varying positions and distances. Meanwhile, scenario video 3 shows several moving human objects with varying positions and distances.

(a) (b)

Source: (Research Results, 2024)

Figure 5. (a) Video Scenario 1 (b) Detection Results of Video 1

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

VOL. 10. NO. 1 AUGUST 2024
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v10i1.5576

(a) (b)

Source: (Research Results, 2024)

Figure 6. (a) Video Scenario 2 (b) Detection Results of Video 2

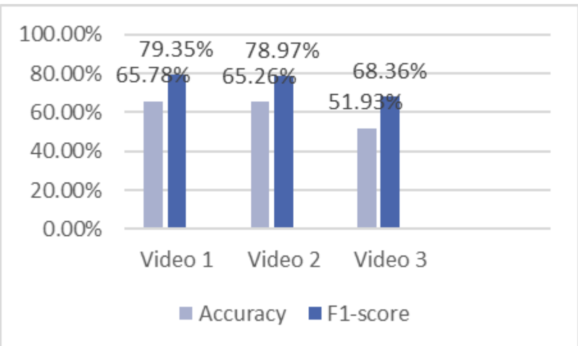
(a) (b)

Source: (Research Results, 2024)

Figure 7. (a) Video Scenario 3 (b) Detection Results of Video 3

Based on Figures 5, 6, and 7, the video scenarios before detection and their detection results are shown. Several human objects have been successfully detected. Detected human objects are marked with a bounding box indicating the object's position, accompanied by class label, confidence score, and distance value for each detection. Human objects detected at a distance of 3 meters or more are displayed with a green bounding box, while those detected at a distance of less than 3 meters are flagged as high risk with a red bounding box.

The system evaluation parameters used include accuracy and F1-score. Accuracy is calculated as the ratio of the number of correct detections to the total expected detections [35], while the F1-score is the harmonic mean of precision and recall [37]. Precision measures the proportion of correct detections produced [36], while recall measures the proportion of correct detections out of all objects that should be detected [35]. The results of human detection testing are presented in Figure 8.

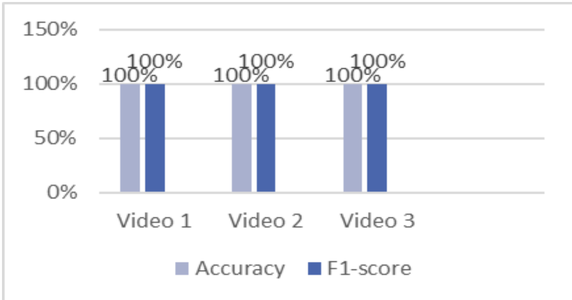


Source: (Research Results, 2024)

Figure 8. Human Detection Test Results

Based on Figure 8, the results of human detection testing on videos 1, 2, and 3 yielded an average accuracy of 60.99% and an F1-score of 75.56%. Many human objects were too small to be clearly detected. In Addition, human objects that are too far from the camera and lack contrast with the background also affect detection accuracy. These detection errors mainly occur in human objects that are small or too far from the camera and have low contrast with the background. This indicates that YOLOv5 has limitations in detecting objects with low resolution or under suboptimal lighting conditions.

In addition to detecting human objects, distance detection from human objects was also performed. The main focus was on detecting high-risk distances to ensure that the system could provide quick, accurate responses and appropriate warnings when humans were within a distance of less than 3 meters. The results of high-risk detection testing are displayed in Figure 9.

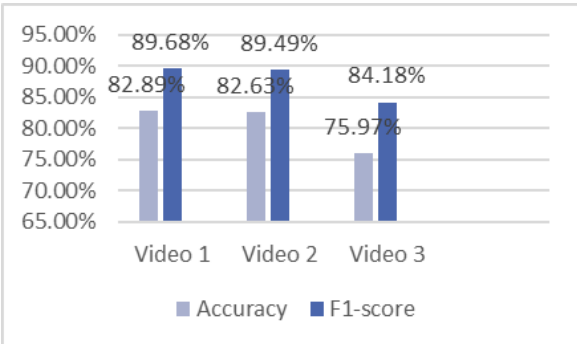


Source: (Research Results, 2024)

Figure 9. High-Risk Detection Test Results

Based on Figure 9, the results of high-risk detection testing on videos 1, 2, and 3 showed an average accuracy of 100% and an F1-score of 100%. These high values indicate that the model performed exceptionally well in achieving the goals of this research.

To assess the overall performance of the model, the evaluation results from the conducted tests are summarized in Figure 10.



Source: (Research Results, 2024)

Figure 10. Detection Evaluation Results

Based on Figure 10, the overall evaluation results of detection on videos 1, 2, and 3 yielded an average accuracy of 80.5% and an F1-score of 87.78%. These high accuracy and F1-score values indicate that the human detection system performs well. Although the accuracy and F1-score have not reached a perfect value of 100% due to several factors, such as the presence of many small human objects, objects that are too far from the camera, and low contrast with the background, these values are sufficiently high to demonstrate the effectiveness of the developed detection system.

The limitations in this study open opportunities for further development. With a dataset comprising only three videos, there is significant potential to expand this research by using more and varied videos to enhance the model's ability to adapt to various real-world situations. Additionally, although the selected videos do not represent all possible field scenarios, this provides a strong foundation for future research that can broaden the scope and improve detection accuracy under various conditions.

In this research, YOLOv5 demonstrates superior performance in detecting human objects in the area behind the excavator compared to previous object detection methods. Studies using Mask R-CNN show high accuracy, but often require longer computation times and greater resources [12]. Faster R-CNN provides good real-time detection [13], but this research shows that YOLOv5 excels in both speed and detection accuracy. Another study that integrates YOLOv4 with a Siamese network for real-time tracking shows effectiveness [14], but YOLOv5 handles background detection and small object issues more effectively to mosaic data augmentation [18]. Additionally, YOLOv5 offers significant improvements in detection speed and accuracy compared to YOLOv3 and other two-stage detection methods like Faster R-CNN [15], [17]. With its excellent performance, YOLOv5 shows strong potential for real-time applications in safety monitoring around heavy equipment. Therefore, YOLOv5 is a better choice due to its accuracy and speed in detecting objects [38].

CONCLUSION

Based on the research results, it can be concluded that the YOLOv5 algorithm demonstrates good performance with an average accuracy of 80.5% and an F1 score of 87.78%, making it suitable for application in human detection systems in the area behind excavators, in line with the research objectives. The implementation of the YOLOv5-based human detection system contributes significantly to the development of safety technology for excavators, especially since YOLOv5 offers advantages in detection speed and accuracy compared to previous detection methods that often require longer computation times and more resources. This contribution aligns with the theory that deep learning-based image processing technology can substantially improve workplace safety by providing more accurate and responsive detection tools. By implementing this system, it is expected to enhance the safety of workers operating excavators, reduce the incidence of work accidents caused by excavators in various sectors, and decrease the risk of injury or death due to collisions with excavators. This system is also capable of providing early warnings to excavator operators accurately, thereby preventing accidents that could cause injury or death and improving operational efficiency by minimizing downtime due to accidents.

This research has limitations that need to be considered by future researchers to improve research outcomes. These limitations are related to program accuracy, detection errors, and data set limitations. The accuracy of the program is greatly influenced by the clarity of the input, where human objects that are too small, too far from the camera, or have low contrast with the background are still difficult to detect clearly. Although the YOLOv5 algorithm performs well in various conditions, detecting objects with low resolution or under less ideal lighting conditions still requires further understanding. Additionally, the dataset used in this research does not fully represent various field situations. To enhance the model's ability to detect human objects in various real-world scenarios, it is recommended to expand the dataset with more and varied videos. Diversity in the dataset will help the model adapt to various situations in the excavator work environment, resulting in more reliable and widely applicable applications. By addressing these limitations, future researchers are expected to produce a more accurate and effective detection system.

As a recommendation for future research, the performance of YOLOv5 can be developed in an object detection system based on the Bird's Eye View concept using cameras mounted on top of the excavator. This system provides a wider, more detailed, and clearer view compared to single-camera systems. Additional benefits of this system include a 360-degree view around the excavator, enabling the detection of objects not visible from the perspective of the rear-mounted camera, multi-directional detection that enhances the ability to detect objects from various directions

VOL. 10. NO. 1 AUGUST 2024
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v10i1.5576

simultaneously, reducing blind spots present in single-camera systems, improving safety by ensuring better detection of workers around the excavator, reducing the risk of accidents, and enhancing response to potential hazards. With proper implementation, this system can detect more objects or people around the excavator from all directions, thus being more effective in enhancing work safety, especially in preventing collisions from various directions.

REFERENCE

[1] *Peraturan Menteri Ketenagakerjaan Republik Indonesia Nomor 8 Tahun 2020 tentang Keselamatan dan Kesehatan Kerja Pesawat Angkat dan Pesawat Angkut*. JDIH BPK, 2020.

[2] R. Anas, "Tergilas Excavator, Seorang Karyawan PT IWIP Dilarikan ke Rumah Sakit," *Beritadetik.id*, Halmahera Tengah, Jul. 05, 2021.

[3] J. Simanjuntak, "Kuli Panggul Terlindas Alat Berat di Pasar Induk Cibitung, Polisi: Masih Hidup," *SINDONEWS.com*, Bekasi, Jul. 07, 2023.

[4] B. Supriadi and P. Kurniati, "Kronologi Pekerja Tewas Terlindas Ekskavator di Tambang Ilegal Jember, 5 Orang Jadi Tersangka," *KOMPAS.com*, Jember, Nov. 10, 2023.

[5] S. F. Putra, T. Rakhmadi, and D. Atmoko, "Analisa Penyebab Blind Spot Pada Truck dan Upaya Mengatasinya," *Bhamada Occupational Health Safety Enironment Journal*, vol. 1, no. 2, pp. 28–35, 2023.

[6] D.-J. Yeom, H.-S. Yoo, J.-S. Kim, and Y. S. Kim, "Development of a Vision-Based Machine Guidance System for Hydraulic Excavators," *Journal of Asian Architecture and Building Engineering*, vol. 22, no. 3, pp. 1564–1581, May 2023, doi: 10.1080/13467581.2022.2090365.

[7] I. Yufahmi, R. Har, and J. Andas, "Analisis Risiko Bahaya dan Upaya Pengendalian Kecelakaan Kerja dengan Metode Hirarki Pengendalian Bahaya pada Area Penambangan Batu Gamping Bukit Karang Putih di PT. Semen Padang, Sumatera Barat," *Jurnal Bina Tambang*, vol. 6, no. 4, 2021, doi: 10.24036/bt.v6i4.114089.

[8] M. Vicky, "Sistem Pendeteksi Objek Pada Area Blind Spot Dump Truck Menggunakan Fuzzy Logic Dengan Metode Sugeno," Universitas Islam Negeri Syarif Hidayatullah Jakarta, 2020.

[9] B. Kim, E.-J. An, S. Kim, K. R. Sri Preethaa, D.-E. Lee, and R. R. Lukacs, "SRGAN-enhanced Unsafe Operation Detection and Classification of Heavy Construction Machinery Using Cascade Learning," *Artif Intell Rev*, vol. 57, no. 8, p. 206, Jul. 2024, doi: 10.1007/s10462-024-10839-7.

[10] M. Park, D. Q. Tran, D. Jung, and S. Park, "Wildfire-Detection Method Using DenseNet and CycleGAN Data Augmentation-Based Remote Camera Imagery," *Remote Sens (Basel)*, vol. 12, no. 22, p. 3715, Nov. 2020, doi: 10.3390/rs12223715.

[11] D. Q. Tran, M. Park, Y. Jeon, J. Bak, and S. Park, "Forest-Fire Response System Using Deep-Learning-Based Approaches With CCTV Images and Weather Data," *IEEE Access*, vol. 10, pp. 66061–66071, 2022, doi: 10.1109/ACCESS.2022.3184707.

[12] O. Angah and A. Y. Chen, "Tracking Multiple Construction Workers Trough Deep Learning and The Gradient Based Method with Re-matching Based on Multi-object Tracking Accuracy," *Autom Constr*, vol. 119, p. 103308, Nov. 2020, doi: 10.1016/j.autcon.2020.103308.

[13] X. Li, T. Hao, F. Li, L. Zhao, and Z. Wang, "Faster R-CNN-LSTM Construction Site Unsafe Behavior Recognition Model," *Applied Sciences*, vol. 13, no. 19, p. 10700, Sep. 2023, doi: 10.3390/app131910700.

[14] H. Son and C. Kim, "Integrated Worker Detection and Tracking for The Safe Operation of Construction Machinery," *Autom Constr*, vol. 126, p. 103670, Jun. 2021, doi: 10.1016/j.autcon.2021.103670.

[15] U. Sirisha, S. P. Praveen, P. N. Srinivasu, P. Barsocchi, and A. K. Bhoi, "Statistical Analysis of Design Aspects of Various YOLO-Based Deep Learning Models for Object Detection," *International Journal of Computational Intelligence Systems*, vol. 16, no. 1, p. 126, Aug. 2023, doi: 10.1007/s44196-023-00302-w.

[16] S. Srivastava, A. V. Divekar, C. Anilkumar, I. Naik, V. Kulkarni, and V. Pattabiraman, "Comparative Analysis of Deep Learning Image Detection Algorithms," *J Big Data*, vol. 8, no. 1, Dec. 2021, doi: 10.1186/s40537-021-00434-w.

[17] T. Zeng, J. Wang, B. Cui, X. Wang, D. Wang, and Y. Zhang, "The Equipment Detection and Localization of Large-Scale Construction Jobsite by Far-Field Construction Surveillance Video Based on Improving YOLOv3 and Grey Wolf Optimizer Improving

Extreme Learning Machine," *Constr Build Mater*, vol. 291, p. 123268, Jul. 2021, doi: 10.1016/j.conbuildmat.2021.123268.

[18] M. Park, D. Q. Tran, J. Bak, and S. Park, "Small and Overlapping Worker Detection at Construction Sites," *Autom Constr*, vol. 151, p. 104856, Jul. 2023, doi: 10.1016/j.autcon.2023.104856.

[19] S. Kim, S. H. Hong, H. Kim, M. Lee, and S. Hwang, "Small Object Detection (SOD) System for Comprehensive Construction Site Safety Monitoring," *Autom Constr*, vol. 156, p. 105103, Dec. 2023, doi: 10.1016/j.autcon.2023.105103.

[20] N. Rajavinu, B. Swaminathan, A. P. Shameer, D. Jagadish, D. Ghosal, and P. B. Chanda, "A Deep Learning-based Approach on Vehicle Detection, Tracking, and Counting System," in *IEEE 5th International Conference on Inventive Computation Technologies, ICICT 2022 - Proceedings*, Institute of Electrical and Electronics Engineers Inc., 2022, pp. 831–836. doi: 10.1109/ICICT54344.2022.9850538.

[21] F. Wu, Z. Cai, S. Fan, R. Song, L. Wang, and W. Cai, "Fish Target Detection in Underwater Blurred Scenes Based on Improved YOLOv5," *IEEE Access*, vol. 11, pp. 122911–122925, 2023, doi: 10.1109/ACCESS.2023.3328940.

[22] D. A. R. Ellen, P. Kristalina, M. Z. S. Hadi, and A. Patriarso, "Effective Searching of Drowning Victims in the River using Deep Learning Method and Underwater Drone," in *IES 2023 - International Electronics Symposium: Unlocking the Potential of Immersive Technology to Live a Better Life, Proceedings of the IEEE*, Institute of Electrical and Electronics Engineers Inc., 2023, pp. 569–574. doi: 10.1109/IES59143.2023.10242589.

[23] Y. Zhou, Z. Wang, G. Xie, J. Liao, W. Huang, and C. Xiao, "A Novel YOLO V5 Framework for Infrared Image Recognition," in *2022 IEEE 5th International Conference on Automation, Electronics and Electrical Engineering, AUTEEE 2022*, Institute of Electrical and Electronics Engineers Inc., 2022, pp. 353–360. doi: 10.1109/AUTEEE56487.2022.9994569.

[24] A. Khalfaoui, A. Badri, and I. El Mourabit, "An Improved YOLOV5 for Real-time Human Detection in Infrared Images," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 32, no. 2, pp. 1078–1085, Nov. 2023, doi: 10.11591/ijeecs.v32.i2.pp1078-1085.

[25] A. Dradrach, J. Konert, and J. Ruminski, "Multimodal Camera for Pedestrian Detection with Deep Learning Models," in *Proceedings of the IEEE International Conference on Industrial Technology*, Institute of Electrical and Electronics Engineers Inc., pp. 1-6, 2023. doi: 10.1109/ICIT58465.2023.10143046.

[26] L. He and H. Wei, "CBAM-YOLOv5: A Promising Network Model for Wear Particle Recognition," *Wirel Commun Mob Comput*, vol. 2023, pp. 1–15, Jun. 2023, doi: 10.1155/2023/2520933.

[27] A. Hossain, M. T. Islam, and A. F. Almutairi, "A Deep Learning Model to Classify and Detect Brain Abnormalities in Portable Microwave Based Imaging System," *Sci Rep*, vol. 12, no. 1, Dec. 2022, doi: 10.1038/s41598-022-10309-6.

[28] S. M. Nasution and M. D. Fussy, "Pedestrian Detection System using YOLOv5 for Advanced Driver Assistance System (ADAS)," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 7, no. 3, pp. 715–721, Jun. 2023, doi: 10.29207/resti.v7i3.4884.

[29] M. S. Sruthi, M. J. Poovathingal, V. N. Nandana, S. Lakshmi, M. Samshad, and V. S. Sudeesh, "YOLOv5 based Open-Source UAV for Human Detection during Search and Rescue (SAR)," in *IEEE 10th International Conference on Advances in Computing and Communications, ICACC 2021*, Institute of Electrical and Electronics Engineers Inc., pp. 1-6, 2021. doi: 10.1109/ICACC-202152719.2021.9708269.

[30] V. K. Paswan and A. Choudhary, "Camera Based Indoor Object Detection and Distance Estimation Framework for Assistive Mobility," in *2022 IEEE International Conference on Service Operations and Logistics, and Informatics, SOLI 2022*, Institute of Electrical and Electronics Engineers Inc., pp. 1-6, 2022. doi: 10.1109/SOLI57430.2022.10294458.

[31] N. Karthik and S. Shenai, "Enabling Object Detection and Distance Calculation in AI based Autonomous Driving System," in *MysuruCon 2022 - 2022 IEEE 2nd Mysore Sub Section International Conference*, Institute of Electrical and Electronics Engineers Inc., pp. 1-6, 2022. doi: 10.1109/MysuruCon55714.2022.9972489.

VOL. 10. NO. 1 AUGUST 2024
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v10i1.5576

[32] C. Duan and S. Luo, "Design of Pedestrian Detection System based on OpenCV," in *Proceedings of the IEEE - 2022 4th International Conference on Artificial Intelligence and Advanced Manufacturing, AIAM 2022*, Institute of Electrical and Electronics Engineers Inc., pp. 256–259, 2022,. doi: 10.1109/AIAM57466.2022.00055.

[33] R. A. Rudiyanto and E. B. Setiawan, "Sentiment Analysis Using Convolutional Neural Network (CNN) and Particle Swarm Optimization on Twitter," *JITK (Jurnal Ilmu Pengetahuan dan Teknologi Komputer)*, vol. 9, no. 2, pp. 188–195, Feb. 2024, doi: 10.33480/jitk.v9i2.5201.

[34] A. F. Cobantoro, F. Masykur, and K. Sussolaikah, "Performance Analysis of Alexnet Convolutional Neural Network (CNN) Architecture with Image Objects of Rice Plant Leaves," *JITK (Jurnal Ilmu Pengetahuan dan Teknologi Komputer)*, vol. 8, no. 2, pp. 111–116, Feb. 2023, doi: 10.33480/jitk.v8i2.4060.

[35] D. Chicco and G. Jurman, "The Advantages of the Matthews Correlation Coefficient (MCC) Over F1 score and Accuracy in Binary Classification Evaluation," *BMC Genomics*, vol. 21, no. 1, Jan. 2020, doi: 10.1186/s12864-019-6413-7.

[36] Ž. Vujović, "Classification Model Evaluation Metrics," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 6, pp. 599–606, 2021, doi: 10.14569/IJACSA.2021.0120670.

[37] O. Nurdiawan, H. Susana, and A. Faqih, "Deep Learning for Polycystic Ovarian Syndrome Classification Using Convolutional Neural Network," *JITK (Jurnal Ilmu Pengetahuan dan Teknologi Komputer)*, vol. 9, no. 2, pp. 218–226, Feb. 2024, doi: 10.33480/jitk.v9i2.4575.

[38] S. Varun and V. M. Bhuvanesh, "Real Time Theft Detection Using YOLOv5 Object Detection Model," in *Proceedings of the IEEE - 2023 3rd International Conference on Innovative Sustainable Computational Technologies, CISCT 2023*, Institute of Electrical and Electronics Engineers Inc., pp. 1-5, 2023. doi: 10.1109/CISCT57197.2023.10351223.