Implementation of Slicing Aided Hyper Inference (SAHI) in YOLOv8 to Counting Oil Palm Trees Using High-Resolution Aerial Imagery Data

Naufal Najiv Zhorif, Rahmat Kenzie Anandyto, Albrizy Ullaya Rusyadi, Edy Irwansyah

Computer Science Department-School of Computer Science, Bina Nusantara University, Jakarta, Indonesia

Abstract—Palm oil is a commodity that contributes significantly to Indonesia's national economic growth, with a total plantation area of 116,000 hectares. By 2023, Indonesia is projected to produce approximately 47 million metric tons of palm oil. One of the major challenges in the manual counting of oil palm trees in a large area of a plantation is the labour-intensive, time-consuming, costly, and dangerous nature of the work in the field. The use of aerial imagery allows for the mapping of large areas with comprehensive data coverage. This study proposes a method of mapping oil palm plantations for the counting of oil palm trees using high-resolution aerial images taken with drones. Furthermore, the use of artificial intelligence (AI) methods and deep learning (DL) with the You Only Look Once (YOLO) model for object detection has demonstrated good accuracy in previous studies. This research will utilize the YOLOv8m object detection model and the slicing method, namely Slicing Hyper Aided Hyper Inference (SAHI), which is anticipated to enhance the precision of object detection models on high-resolution aerial imagery. The study concluded that the use of the SAHI slicing method can significantly enhance the accuracy of the model, as evidenced by a Mean Absolute Percentage Error (MAPE) value of 0.01758 on aerial imagery equivalent to 73.2 hectares, with a detection time of 5 minutes and 45 seconds.

Keywords—Oil palm tree; YOLOv8; SAHI; aerial imagery; tree counting

I. INTRODUCTION

Palm oil ranks as one of the most extensively used edible oils worldwide, with global production reaching 79.31 million metric tons in 2023. This represents a 2% increase from the previous year, when 78 million metric tons were produced. In 2023, Indonesia stands as the foremost producer of palm oil, with an output of 47 million metric tons, followed by Malaysia with 19 million metric tons [1]. To achieve the production figures, Industrial plantations in Indonesia have grown by 116,000 hectares in 2023 [2]. Palm oil represents a significant contributor to Indonesia’s national economic growth, serving as a cornerstone of the country’s commodity exports. Nevertheless, the monitoring of oil palm plantations remains a significant challenge. The manual monitoring process is labor-intensive, time-consuming, costly, and poses inherent risks to workers on oil palm plantations. To address these challenges, aerial imagery technology can be employed to supplement and enhance the limitations of manual monitoring methods, offering more accurate and efficient results while reducing the risks faced by workers. The utilization of aerial imagery technology facilitates the monitoring of plant health, the mapping of oil palm plantation areas, and the detection of oil palm plants [3].

However, the data complexity of aerial imagery necessitates the use of a robust analytical methodology. To analyze the data, artificial intelligence (AI) technology can be employed through techniques such as machine learning (ML) and deep learning (DL) to produce accurate aerial image analysis [4]. When employing AI for the detection and enumeration of oil palm trees, the algorithmic process necessarily entails a learning phase to facilitate the identification of these trees. Subsequently, the resulting detection data can be utilized to ascertain the total number of oil palm trees within the designated area.

One of the most widely used deep learning algorithms for object detection is the convolutional neural network (CNN). This has led to the development of several derivatives, including the region-based convolutional neural network (R-CNN), the faster region-based convolutional neural network (Faster R-CNN), and the mask region-based convolutional neural network (Mask R-CNN) [5]. The CNN algorithm has previously been employed in research to detect and enumerate oil palm trees in high-resolution remote sensing images [6]. Nevertheless, in the context of real-time detection, CNN’s performance still needs improvement, even in the advanced development of CNN, namely Faster R-CNN [7]. The You Only Look Once (YOLO) algorithm exhibits superior real-time performance.

The YOLO object detection deep learning model has been designed to achieve high-performance levels in real-time detection. However, despite this, its accuracy remains below that of the Faster R-CNN model [8]. Furthermore, the YOLO model has been observed to experience difficulties in the detection of smaller objects [7]. Nevertheless, the YOLO model does possess an advantage in terms of its speed and performance in real-time object detection.

This research will implement artificial intelligence, specifically the deep learning YOLO algorithm, as a model for palm oil tree detection and Sliced Aided Hyper Inference (SAHI) [9] to enhance the precision of YOLO detection on high-resolution aerial imagery. The data utilized in this study was obtained from aerial images captured using unmanned aerial vehicles (UAVs) in the oil palm plantation region of North Sumatra, Indonesia.

The structure of this paper consists of previous work, methodology, results, discussion, and conclusion. In the previous work section, discussed previous research that is
relevant to this research and what the focus of this research is. Methodology discusses the methods that are used in this research, starting from data collection, pre-processing, and model development to the model evaluation method used. The result section discusses the results of the training model, model detection by comparing the results without the SAHI method and by using the SAHI method, and the results of the model evaluation. The discussion part discusses the results of the research that has been done and compares it with research that has been done before. Besides that, it will also discuss the contributions obtained from this research, its shortcomings, and suggestions for further research. Conclusion will discuss the conclusions of the research that has been done.

II. PREVIOUS WORK

Previous researchers have employed the deep learning YOLO model for oil palm tree detection with a variety of datasets (see Table I). According to a study by [10], the efficacy of YOLOv3 for oil palm tree detection using remote sensing data was evaluated, and the result was an evaluation value of 0.057627 based on the MAPE metric. In a related study, the authors employed YOLOv3, v4, and v5m to detect oil palm trees using aerial imagery data collected from VTOL drones in oil palm plantations in Jambi province, Indonesia. The F1-Score evaluation value for YOLOv3 was 97.28%, for YOLOv4, it was 97.74%, and for YOLOv5m, it was 94.94%. The research conducted by [12] utilizing the Deep Learning Faster R-CNN algorithm for the detection, counting, and geolocation of palm trees achieved an evaluation value of 94% in precision, 84% in recall, and 83% in AP IoU values in plantation areas in the Khajr region of Saudi Arabia. A study conducted by study [13] utilizing YOLOv8 and aerial imagery data yielded an overall accuracy value of 98.50% in the oil palm plantation area in West Kalimantan province, Indonesia. The study in [14] and [15] conducted a similar study using YOLOv5 to perform detection and classification. The classification was divided into five categories: healthy, smallish, yellowish, mismanaged, and dead palms. The F1-Score evaluation values ranged from 0.82 to 0.895. In a recent study [16], the modified YOLOv3n algorithm was employed to perform real-time detection in oil palm plantation areas, resulting in an F1-Score of 0.91 and a mAP of 97.20.

This research project will focus on the detection and counting of oil palm trees through the implementation of artificial intelligence, precisely by utilizing the deep learning algorithm YOLO as a model for oil palm tree detection and Slicing Aided Hyper Inference (SAHI) [9] to enhance the accuracy of YOLO detection on high-resolution aerial imagery. The data utilized in this study was obtained from aerial imagery captured using a drone in the oil palm plantation industry area in North Sumatra, Indonesia.

III. METHODOLOGY

The research comprises several stages, including data collection, using drones to obtain images for datasets, and model evaluation. The data is then pre-processed, dividing the images captured by the drone into four different categories: training data, validation data, test data, and evaluation data. This allows for the model to be tested. The next stage is model development, which involves setting up the YOLOv8m [17] training model and YOLOv8m training model and setting up Slicing Aided Hyper Inference (SAHI) [18]. Finally, the model is evaluated using images that are distinct from those used in the dataset. The research process is illustrated in the flowchart “Fig. 1”.

### A. Data Collecting and Research Area

The data collection location for this research was an oil palm plantation situated in the Gunung Bayu area, Afdeling IV, blocks 13L, 13M, 12AK (Dataset), 14AC, 14AE, 14AF (Evaluation), North Sumatra, Indonesia “Fig. 2”.

The dataset was collected using a Trinity F90+ drone “Fig. 3” flying 200 meters above ground level, which captured RGB images. Table II shows drone specification.

### B. Data Pre-processing

The recently acquired data will undergo a preliminary processing phase before integrating into the model’s datasets and

---

**TABLE I. PREVIOUS RESEARCH**

<table>
<thead>
<tr>
<th>No</th>
<th>Topics</th>
<th>Author and Year</th>
<th>Method</th>
<th>Evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Palm Oil Tree Counting with Remote Sensing Imagery</td>
<td>Mukhes Sri Muna et al., 2022</td>
<td>YOLOv3</td>
<td>5.76% (MAPE)</td>
</tr>
<tr>
<td>2.</td>
<td>Oil Palm Trees Detection with High-Resolution Remote Sensing Image</td>
<td>Hery Wibowo et al., 2022</td>
<td>YOLOv3, YOLOv4, YOLOv5m</td>
<td>97.28% (v3), 97.74% (v4), 94.94% (v5m), (F1-Score)</td>
</tr>
<tr>
<td>3.</td>
<td>Palm Tree Counting and Geolocation</td>
<td>Adel Ammar et al., 2021</td>
<td>Faster R-CNN</td>
<td>94% (Precision), 84% (Recall), 83% (AP IoU),</td>
</tr>
<tr>
<td>4.</td>
<td>Oil Palm Trees Detection with High-Resolution Aerial Image Data</td>
<td>Wardana et al., 2023</td>
<td>YOLOv8</td>
<td>98.50% (Overall Accuracy)</td>
</tr>
<tr>
<td>5.</td>
<td>Oil Palm Trees Detection with YOLO-V5</td>
<td>Desta Sandya Prasвитa et al, 2023</td>
<td>YOLOv5x, YOLOv5m, YOLOv5l, YOLOv5x</td>
<td>0.82 (v5x), 0.84(v5m), 0.85 (v5l), 0.86 (v5x) (Average F1-Score)</td>
</tr>
<tr>
<td>6.</td>
<td>Monitoring Oil Palm Tree Health with YOLOv5</td>
<td>Nuwara et al., 2022</td>
<td>YOLOv5</td>
<td>0.895 (F1-Score)</td>
</tr>
<tr>
<td>7.</td>
<td>Object Detection in Oil Palm Plantation using a Hybrid Feature Extractor of YOLO-based Model</td>
<td>Junos et al., 2022</td>
<td>YOLOv3n</td>
<td>97.20% (mAP), 0.91 (F1-Score)</td>
</tr>
</tbody>
</table>
evaluation data. The following describes the preprocessing stages that will be employed.

1) The images captured by the drone will undergo a stitching process utilizing the Agisoft Metashape software. The outcome of this process is a comprehensive representation of each garden block in the form of a GeoTIFF file.

2) GeoTIFF images will be converted into .png format using the ArcGIS Pro software.

3) The image will be sliced using Adobe Photoshop. The slicing process employs a guide layout comprising 13 columns and 13 rows. Each sliced image has a size of 1399 x 1392 pixels “Fig. 4”.

4) The annotation process is conducted using Roboflow, with a total of 200 images utilized as a dataset comprising 111 images for training, 50 images for validation, and 39 images for testing. When performing annotation, augmentation techniques such as flip (horizontal, vertical), and 90° rotate (clockwise, counterclockwise, upside down) are employed, resulting in an increased training dataset size to 516 images “Fig. 5”.

Fig. 1. Research flowchart.

Table II. Drone Specification Used for Data Collection [19]

<table>
<thead>
<tr>
<th>Specification</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transmitter Frequency</td>
<td>2.4 GHz</td>
</tr>
<tr>
<td>Command and Control Range</td>
<td>5 – 7.5 Km</td>
</tr>
<tr>
<td>Max Flight Time</td>
<td>90 Minutes</td>
</tr>
<tr>
<td>Camera Sensor</td>
<td>Sony RX1R II 42.4 MP</td>
</tr>
<tr>
<td>Max Range, Area</td>
<td>100 Km, 700 hectares</td>
</tr>
</tbody>
</table>

Fig. 2. Research area is located in North Sumatra, Indonesia, at 99.3620323°E and 3.1182583°N. The blue rectangle indicates the extent of the farm area included in the dataset, while the red rectangle represents the farm area utilized for model evaluation purposes.

Fig. 3. Trinity F90+ drone.

Fig. 4. The slicing process for dataset images results in an image with a size of 1399 x 1392 pixels for each image that has been sliced.
Fig. 5. The annotation process, conducted using Roboflow, is confined to the middle of the crown of the oil palm tree. This is done to ensure that the bounding box detection results do not overlap.

C. Model Development

The YOLOv8m model was trained using Google Collaboratory, which was equipped with 85 GB of memory and a 40 GB NVIDIA A100 GPU. The training process utilized an image size of 800 pixels, and the other parameters were set to their default values (see Fig. 6).

The next step is to implement SAHI (Fig. 7) by utilizing the weight of the training model results that have previously been carried out. This study employs a slicing height of 3000px and a slicing width of 3000px, a model confidence threshold of 0.55, and an overlap height ratio and overlap width ratio of 0.2 (default setting).

D. Model Evaluation Method

To evaluate the performance of the detection model, this study uses the Mean Absolute Percentage Error (MAPE). This is achieved by comparing the number of detections made by the model with the actual number of objects in the field.

\[
MAPE = \frac{1}{n} \sum_{t=1}^{n} \left| \frac{A_t - F_t}{A_t} \right|
\]  \hspace{1cm} (1)

IV. RESULT AND DISCUSSION

A. Training Model Result

The results of the YOLOv8m training model, with using an image size 800 pixel, indicate a Recall value of 0.899, a Precision value of 0.932, and an F1-Score value of 0.916. The training model was executed for 78 epochs, with a total runtime of 9 minutes and 39 seconds.

The results of training and validation models “Fig. 8”, show that the Training loss moves down consistently.

With a drastic decrease at epoch two from a value of 2.013 to 1.339, a slight spike in loss occurs at epoch three and then back down. In Validation loss, there is a very drastic increase in loss at epoch 3, with a loss value of 2,336, where the previous value was 1,327. The loss then returned to 1,534, after which the validation loss experienced significant fluctuations until epoch 53, after which the validation loss became more stable, but there was no significant improvement. Although the movement of validation loss does not mirror that of training loss, as illustrated in “Fig. 8”, the training model does not shows signs of overfitting.

“Fig. 9” illustrates the values of precision, recall, and F1 score.

This illustrates a notable decline in these values at the epoch tree, from 0.733 of precision, 0.993 of recall, and 0.844 of F1-score to 0.150 of precision, 0.213 of recall, and 0.176 of F1-score. However, there was a subsequent surge in these values,
reaching 0.829 of precision, 0.868 of recall, and 0.848 of F1-score. A further decline was observed at epoch 16, although this was not statistically significant. The values decreased from 0.890 of precision, 0.976 of recall, and 0.931 of F1-score to 0.865 of precision, 0.908 of recall, and 0.886 of F1-score.

B. Evaluation Model Result

For detection without the SAHI method “Fig. 10”, the detection results on an area of 73.2 hectares with a confidence value of 0.55, which is the same confidence value used for detection with SAHI, show that the model does not detect any trees. At a confidence value of 0.1, there is a bounding box that does not show good detection results. A test using a confidence value of 0.55 and a plantation area of 1 hectare showed good detection results.

To evaluate the model, data in the form of images with a size of 36336 pixels x 27084 pixels are utilized. The image encompasses three plantation blocks (14AC, 14AE, and 14AF) with an area of 73.2 hectares “Fig. 10”.

The results of the SAHI method “Fig. 10” indicate that the model has identified 9,498 oil palm trees, whereas the actual number of oil palm trees is 9,668. Block 14AC has contains 3,154 trees, Block 14AF has contains 3,278 trees, and Block 14AE has contains 3,236 trees, for a total of three blocks containing 9,668 oil palm trees. Fig. 11 shows palm oil plantation.

To calculate the accuracy of the oil palm tree detection result by the model, the Mean Absolute Percentage Error (MAPE) is used as a measure of model accuracy, according to the results of the MAPE calculation Table III.

| TABLE III. MAPE CALCULATION RESULT |
|-------------------------------|-----------------|-----------------|-----------------|
| Confidence | Without SAHI | With SAHI |
| Area (hectares) | 73.2 | 73.2 | 1 |
| Model Detection | 0 | 0 | 147 |
| Actual Trees | 9668 | 9668 | 151 |
| MAPE | 0.01758 | 0.02649 |

The MAPE result shows the model error in the number of oil palm tree detection results compared to the actual number. The MAPE value is close to 0, so the model has good accuracy [21].

For detection that does not use SAHI on an area of 73.2 hectares using a confidence value of 0.1 and 0.55, get a MAPE value of 1 or 100% because no palm trees are detected by the object. However, in an area of 1 hectare with a confidence value of 0.55, a MAPE value of 0.02649 or 2.64% is obtained, which is a good value in an area of 1 hectare without using SAHI.

And in an area of 73.2 hectares with using SAHI, a MAPE value of 0.01758 or 1.75% is obtained, which is a very good value for detection in an area of 73.2 hectares.

C. Discussion

YOLOv8m produces high accuracy with a Mean Absolute Percentage Error (MAPE) value of 0.0175 or 1.75% on aerial imagery with a resolution of 36336 x 27084 pixels, equivalent to 73.2 ha. This area is six times larger than the previous study [11] and 73 times larger than the study [13].

However, the results of this study are limited by the longer detection times observed when compared to previous research, such as that presented in study [11] and [13]. In research [11], the slowest detection took 45 seconds (YOLOv4) and the fastest 21 seconds (YOLOv5m) on an area of 12 ha. In this study, with the same large area of 12 hectares, the results took 45 seconds, but when compared with YOLOv5m, it was 14 seconds longer. In research [13], the time taken to detect oil palm trees on an area of 1 hectare was found to be 16.1 ms, which is considerably faster than the results of this study, where the same task on an area of 1 hectare took 6 seconds. However, this research has been able to address one of YOLO’s weaknesses, namely the difficulty of detecting small objects, as demonstrated in “Fig. 9” and “Fig. 10”.

Using datasets that need more diversity leads to suboptimal detection results on trees exhibiting unhealthy, dry, or dead characteristics. To enhance future research, it is recommended
to utilize a more diverse and comprehensive set of datasets, thereby enabling the model to effectively identify oil palm trees with unhealthy, dry, or dead tree characteristics. Additionally, performing hyperparameter tuning during the training process can further optimize the model training outcomes.

V. CONCLUSION

In this study, the detection and counting of palm oil trees on high-resolution aerial images has been carried out using the Deep Learning YOLOv8sm method in conjunction with the Slicing Aided Hyper Inference (SAHI) method. The results of testing the model on an image of an oil palm plantation area with a resolution of 36336 pixels x 27084 pixels, which corresponds to an area of 73.2 hectares and comprises three gardens, are presented below. In the blocks (14AC, 14AE, and 14AF), there are 9668 oil palm trees in the garden area, with the model successfully detecting as many as 9498 trees. The model's high level of accuracy can be attributed to the use of a small dataset and minimal tuning. This results in a MAPE value of 1.75% and a processing time of 5 minutes and 45 seconds for an image of 73.2 hectares. The combination of YOLOv8 and SAHI is highly recommended for object detection on aerial and satellite images due to its high accuracy on high-resolution aerial image.

ACKNOWLEDGMENT

This work is supported by Bina Nusantara University as a part of Bina Nusantara University’s BINUS International Research “Mapping Oil Palm Plantation Area with Deep Learning Semantic Segmentation Approach” with contract number: 097/VRRTT/VII/2024 and contract date 02 July 2024.

REFERENCES