**ELKOMIKA: Jurnal Teknik Energi Elektrik, Teknik Telekomunikasi, & Teknik Elektronika**

ISSN(p): 2338-8323 | ISSN(e): 2459-9638 | Vol. 14 | No. 3 | Halaman 288 – 302
DOI : http://dx.doi.org/10.26760/elkomika.v14i3.288 July 2026

# Comparative Analysis of YOLOv8 and YOLOv11 for Wildlife Detection in Natural Habitats

**FELIX CORPUTTY[1,2], AHMAD SULTHON JAUHARI[1,4], DARIELE ZEBADA SANUWU GEA[1,4], IVAN FEBRIANTO LALO[1,4], RIFKI RAHMAN NUR IKHSAN[1,3], ADE ADITYA RAMADHA[1,2]**

[1]School of Electrical Engineering, Telkom University, Bandung, Indonesia
[2]Center of Excellence for Telecom Infra Project, Telkom University, Bandung, Indonesia
[3]Center of Excellence for Sustainable Energi and Climate Energi, Telkom University, Bandung, Indonesia
[4]Image Processing and Vision Laboratory, Telkom University, Bandung, Indonesia
Email: felixcorputtyfc@telkomuniversity.ac.id

***Article History:***
*Received*
*2 March 2026*
*Revised*
*5 May 2026*
*Accepted*
*21 May 2026*

**ABSTRAK**

*Pemantauan satwa liar di habitat alami tetap menghadapi tantangan besar akibat occlusion, kamuflase, dan variasi pencahayaan yang membatasi kinerja sistem deteksi objek konvensional. Meskipun model berbasis YOLO terbaru menunjukkan potensi yang kuat, perbandingan sistematis antar arsitektur terbaru di lingkungan ekologi nyata masih terbatas. Penelitian ini mengevaluasi YOLOv8 dan YOLOv11 menggunakan 6.165 citra dari 14 kelas satwa yang dikumpulkan di Taman Buru Gunung Masigit Kareumbi, dengan penerapan augmentasi data untuk meningkatkan robustnes. Kinerja dievaluasi menggunakan pendekatan komparatif kuantitatif yang didasarkan pada presisi, recall, mAP 0.5, dan mAP 0.5:0.95. Hasil menunjukkan bahwa YOLOv11 mencapai akurasi deteksi lebih tinggi dengan nilai mAP 0.5:0.95 sebesar 0.894.*

***Kata kunci**: Object Detection, Wildlife Detection, YOLOv8, YOLOv11.*

**ABSTRACT**

*Wildlife monitoring in natural habitats remains difficult due to occlusion, camouflage, and varying illumination, which limit the performance of conventional object detection systems. Although recent YOLO-based models have shown strong potential, systematic comparisons between newer architectures in real ecological environments are still limited. This study evaluates YOLOv8 and YOLOv11 using 6,165 images from 14 wildlife classes collected at Taman Buru Gunung Masigit Kareumbi, with data augmentation applied to improve robustness. Performance is evaluated using a quantitative comparative approach based on precision, recall, mAP 0.5, and mAP 0.5:0.95. The results indicate that YOLOv11 achieves higher detection accuracy with an mAP 0.5:0.95 of 0.894.*

***Keywords**: Object Detection, Wildlife Detection, YOLOv8, YOLOv11.*

## 1. INTRODUCTION

Biodiversity is one of the main indicators of ecosystem balance on Earth. However, increasing human activities such as deforestation, hunting, and climate change have accelerated the decline of wildlife populations in various regions **(Roy et al., 2023)(Yang et al., 2023)**. Conventional wildlife conservation efforts still rely heavily on manual observation and the use of camera traps, which are limited in terms of coverage, efficiency, and accuracy **(Ma & Yang, 2022)(Zhang & Cai, 2023)**. These limitations have driven the need for technology-based automated monitoring systems that are capable of operating in real time, adaptively, and efficiently in natural environments **(Jiang et al., 2022)**. Considerable advancements in deep learning and computer vision throughout the past decade have brought about a paradigm shift in the strategies employed for wildlife conservation. The You Only Look Once (YOLO) algorithm has gained widespread recognition in the field of object detection, primarily because of its single-pass detection mechanism that offers an effective balance between speed and accuracy **(Jiang et al., 2022)(Naidu et al., 2025)**. Early versions of YOLO, such as YOLOv3 and YOLOv4, were capable of recognizing general objects but faced challenges in detecting small, camouflaged wildlife or animals overlapping with natural vegetation **(May et al., 2025)**.

To address these limitations, numerous studies have focused on enhancing YOLO architectures to better suit ecological contexts. Roy et al. **(Roy et al., 2023)** introduced WilDect-YOLO, a model designed to enhance the detection accuracy of rare species by implementing structural modifications to the YOLOv5 architecture. Meanwhile, Yang et al. **(Yang et al., 2023)** introduced an enhanced YOLOv5s model for forest wildlife detection with improved inference efficiency. Aerial data–based approaches have also been explored, such as CE-RetinaNet proposed by Zhang and Cai **(Zhang & Cai, 2023)**, which incorporates channel enhancement to detect wildlife from infrared UAV imagery, and YOLO-Animal by Ma and Yang **(Ma & Yang, 2022)**, which optimizes multi-class detection in natural environments. In addition, studies by Roca et al. **(Roca et al., 2025)** and Naidu et al. **(Naidu et al., 2025)** employed drones as the primary platform for long-range wildlife monitoring, aiming to increase detection speed while minimizing disturbance to animal behavior. May et al. **(May et al., 2025)** highlighted data annotation efficiency as a critical issue in training large-scale wildlife detection models. This perspective is further supported by the work of Jiang and Wu **(Jiang & Wu, 2024)**, who combined an Extended Kalman Filter with YOLOv8 to enhance the tracking of moving wildlife in complex environments. The development of YOLOv8 represents a major leap in modern detection architectures through the integration of an anchor-free head, decoupled detection layers, and a C2f backbone **(Chen et al., 2024)(John & Issac, 2025)**. This model has been shown to be more efficient for edge-AI–based systems, as demonstrated in studies by Chappidi and Sundaram **(Chappidi & Sundaram, 2024)**. Furthermore, Prabu et al. **(Prabu et al., 2026)** combined YOLOv8 with a 3D Transformer and Adaptive GRU for nighttime detection, while Feng et al. **(Feng et al., 2024)** demonstrated YOLOv8's effectiveness in detecting small objects from aerial imagery. Subsequently, Karanam et al. **(Karanam et al., 2026)** demonstrated the effectiveness of YOLOv8 in recognizing bird species for habitat conservation, highlighting the potential application of the latest YOLO model for biodiversity protection.

The subsequent evolution, YOLOv9–YOLOv11, introduced architectural updates including transformer-based backbones, hybrid attention mechanisms, and multi-modal capabilities. Li et al. **(Li et al., 2025)** presented YOLOv11-OPNet, a bimodal architecture engineered for the automated detection of fishing vessels in maritime environments, while Fang et al. **(Fang et al., 2025)** designed Birds-YOLO based on YOLOv11 for monitoring bird populations in Dongting Lake. Zhao et al. **(Zhao et al., 2025)** applied an Enhanced YOLOv11 to autonomously detect Ochotona curzoniae burrows using UAVs, achieving an accuracy improvement of more than 6% compared to YOLOv8. Studies by Sharma et al. **(Sharma et al., 2024)** and Alkhammash **(Alkhammash, 2025)** further reinforced these findings, showing that YOLOv11 outperforms YOLOv8, YOLOv9, and YOLOv10 in both accuracy and computational efficiency. Beyond terrestrial conservation, YOLO has also been widely applied in aquatic environments. Jiang et al. **(Jiang et al., 2025)** introduced Mobile-YOLO for lightweight aquatic organism detection, while Lawal et al**. (Lawal et al., 2025)** developed YOLO-based MADNet for marine animal detection in coral reef ecosystems. On the other hand, Basit Mughal et al. **(Mughal et al., 2025)** emphasized the potential of integrating edge-AI for real-time wildlife detection, enabling deployment on low-power field devices. This aligns with the study by Adhikari et al. **(Adhikari et al., 2025)**, who systematically evaluated the capability of YOLO models running on edge hardware specifically for deer detection tasks. Research by Gokulapriya and Princy Suganthi **(Gokulapriya & Princy Suganthi, 2025)** also confirmed the effectiveness of YOLOv8 for animal surveillance applications in rural areas with low latency.

While YOLO-based deep learning frameworks have seen extensive application in wildlife detection tasks, the existing body of research continues to present certain notable constraints that warrant further investigation. Most previous research evaluates model performance under relatively controlled conditions, which do not fully represent the visual complexity of natural habitats characterized by camouflage, dense vegetation, occlusion, varying illumination, and dynamic animal movement. Furthermore, prior studies generally focus on improving individual model performance without conducting a comprehensive comparative analysis of the trade-offs among detection accuracy, robustness to visual disturbances, and computational efficiency—factors that are critical for long-term wildlife monitoring systems deployed in real-world field conditions.

YOLOv8 and YOLOv11 represent two distinct architectural approaches, with YOLOv8 emphasizing model efficiency and inference speed, while YOLOv11 is designed to enhance spatial perception and robustness to visual complexity **(O'Reilly et al., 2025)**. Nevertheless, a systematic comparison of the performance of these two models specifically in the context of wildlife detection within natural habitats has yet to be conducted. This research gap results in uncertainty regarding the selection of the most suitable model to support robust, efficient, and sustainable wildlife conservation detection systems. Therefore, the results of this study are expected to contribute to the development of automated wildlife monitoring systems that are robust, efficient, and sustainable, thereby supporting global conservation initiatives in addressing the biodiversity crisis **(Charanya et al., 2025)(Zhong et al., 2022)**.

## 2. METHOD

### 2.1 Research Framework

This study adopts a quantitative approach based on computational experiments with the objective of comparing the performance of the YOLOv8 and YOLOv11 architectures in detecting wildlife in natural habitat imagery. The research workflow is illustrated in Figure 1 and consists of five main stages: (1) Wildlife Dataset Collection, (2) Wildlife Dataset Preprocessing, (3) Model Construction and Training for Wildlife Detection, (4) Evaluation of the Trained Models, and (5) Comparative Analysis of Detection Results. This experimental approach is designed to examine the performance differences between the YOLOv8 and YOLOv11 object detection models in identifying animals in the wild. The study is also intended to determine which model is more effective and superior in detecting wildlife in open natural environments with high accuracy. The research hypothesizes that YOLOv11 exhibits better detection performance than YOLOv8 in terms of accuracy. The YOLOv8 and YOLOv11 models developed in this study are trained using a wildlife dataset consisting of multiple animal species. The dataset was independently collected by Wanadri as the managing organization of Taman Buru Gunung Masigit Kareumbi.

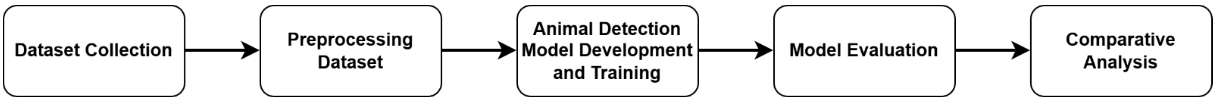


**Figure 1. Research Workflow**

## 2.2 Dataset and Data Preprocessing

Wanadri, the organization responsible for overseeing Taman Buru Gunung Masigit Kareumbi, independently compiled and provided the dataset employed in this study. Examples of the dataset captured by Wanadri using camera traps are shown in Figure 2. The dataset consists of 14 wildlife classes found in Taman Buru Gunung Masigit Kareumbi, namely *Hystrix javanica*, *Lariscus insignis*, *Macaca fascicularis*, *Martes flavigula*, *Melogale orientalis*, *Muntiacus muntjak*, *Panthera pardus melas*, *Paradoxurus hermaphroditus*, *Presbytis comata*, *Prionailurus javanensis*, *Sus scrofa*, *Trachypithecus auratus*, *Urva javanica*, and *Viverricula indica*. Table 1 shows the distribution of the dataset before and after augmentation.


**Figure 2. Sample Images of Dataset – Panthera Pardus Melas (left) and Sus Scrofa (right)**

**Table 1. Dataset**

| Class Name | Original Images | Augmented Images |
|---|---|---|
| Hystrix Javanica | 194 | 468 |
| Lariscus Insignis | 208 | 500 |
| Macaca Fascicularis | 156 | 378 |
| Martes Flavigula | 166 | 448 |
| Melogale Orientalis | 165 | 405 |
| Muntiacus Muntjak | 168 | 414 |
| Panthera Pardus Melas | 443 | 443 |
| Paradoxurus Hermaphoditus | 203 | 499 |
| Presbytis Comata | 135 | 333 |
| Prionailurus Javanensis | 214 | 514 |
| Sus Scrofa | 191 | 467 |
| Trachypithecus Auratus | 254 | 478 |
| Urva Javanica | 177 | 451 |
| Viverricula Indica | 153 | 367 |
| **Total** | **2827** | **6165** |

Data augmentation was applied to address class imbalance and improve model generalization. This study employed an offline augmentation strategy, where additional samples were explicitly generated and added to the dataset. The augmentation process was conducted proportionally across classes based on their initial data distribution. Classes with fewer samples were augmented more extensively to reduce imbalance, while dominant classes were minimally augmented to avoid over-representation. The applied transformations include horizontal flipping, small-angle rotation (−1° to +1°), brightness adjustment (−15% to +15%), saturation variation (−25% to +25%), Gaussian blur up to 2.5 pixels, and noise injection up to 0.1%. As a result, the dataset size increased from 2827 images to 6165 images after augmentation. This approach improves data balance and enhances the model's robustness against real-world environmental variations such as lighting changes, motion blur, and occlusion. The augmentation process was intentionally designed to avoid excessive duplication in dominant classes such as Panthera pardus melas, while significantly increasing minority class samples to achieve a more balanced distribution.

## 2.3 Model Architecture

### 2.3.1 YOLOv8

The YOLOv8 architecture is an advanced development of YOLOv5 and YOLOv7 released by Ultralytics, with a primary focus on computational efficiency and improved detection accuracy through an anchor-free approach and a decoupled head design. Unlike previous versions that relied on anchor-based bounding boxes, YOLOv8 adopts an anchor-free detection system that directly predicts object center positions and bounding box dimensions without dependence on predefined anchor boxes **(Chen et al., 2024)(John & Issac, 2025)**. This method streamlines the training process while strengthening the model's capacity to adapt to objects with different scales and geometries. Its advantages are particularly evident in outdoor ecosystems, where fluctuating illumination and dense vegetation create complex visual conditions typical of wildlife habitats **(Roca et al., 2025)**.

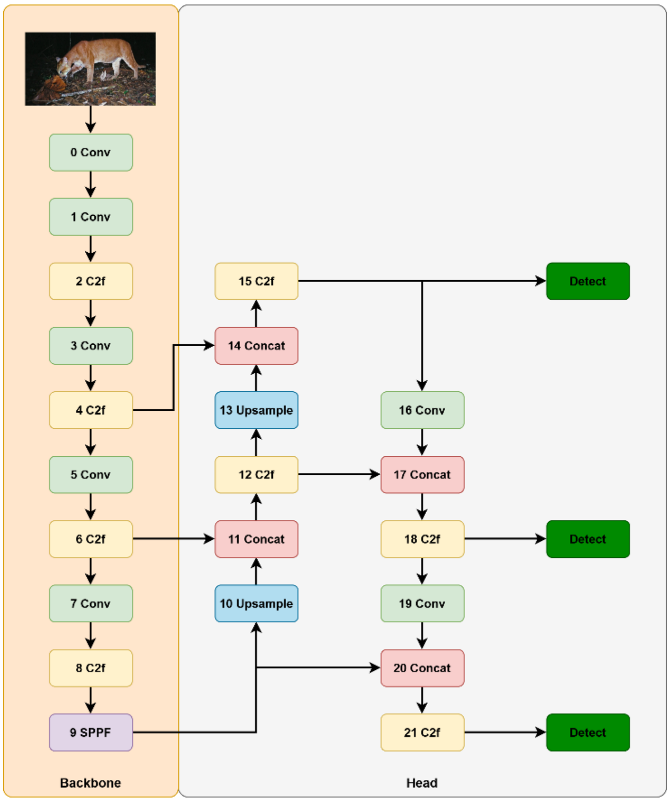


**Figure 3. YOLOv8 Architecture**

From an architectural standpoint, YOLOv8 is built upon three fundamental modules, namely the backbone, neck, and detection head. YOLOv8 uses the same backbone as YOLOv5, namely

CSPDarknet53, with several modifications in which the CSP layers are replaced by C2f (Cross Stage Partial Fusion) modules **(Corputty et al., 2024)** to accelerate gradient flow and reduce feature redundancy across convolutional layers. The integration of C2f and Spatial Pyramid Pooling–Fast (SPPF) modules improves the model's capacity to capture multi-scale features, which are particularly useful for identifying small wildlife objects, including birds and other small-sized animals. In YOLOv8, the neck component utilizes an integration of Feature Pyramid Network (FPN) and Path Aggregation Network (PAN) mechanisms to consolidate features extracted across different resolution scales. PANet adopts an enhanced FPN structure that includes both bottom-up and top-down pathways, enabling better utilization of low-level features and improved object localization accuracy **(Corputty et al., 2022)**. This design allows high-precision detection of small and distant objects. Furthermore, the decoupled detection strategy accelerates training convergence and improves accuracy, particularly in complex scenarios **(Prabu et al., 2026)(Sharma et al., 2024)**. The overall YOLOv8 architecture is illustrated in Figure 3.

### 2.3.2 YOLOv11

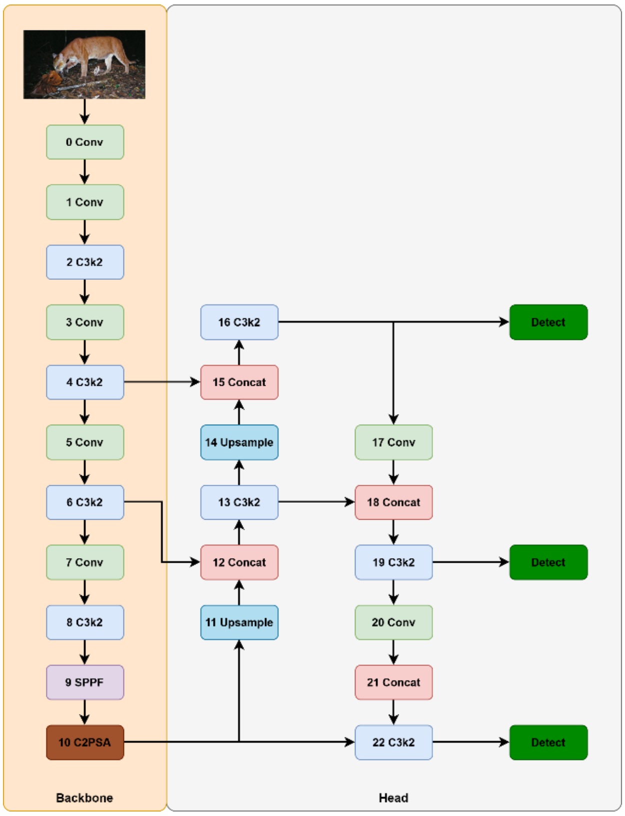


**Figure 4. YOLOv11 Architecture**

The YOLOv11 architecture represents the latest evolution of the YOLO family, introducing fundamental changes in network design by integrating a transformer-based backbone, multi-scale attention mechanisms, and support for dual-modality inputs (RGB and thermal). Unlike YOLOv8, which relies on a purely convolutional structure, YOLOv11 incorporates a CSP-Transformer Encoder in the backbone, enabling the model to capture global spatial context more effectively **(Fang et al., 2025)(Zhao et al., 2025)**. This approach helps the model recognize camouflaged wildlife or animals with textures similar to their surrounding environment, which have traditionally been difficult to detect using conventional CNN-based models **(Zhu et al., 2025)**.

In the neck component, YOLOv11 employs a Bi-directional Feature Pyramid Network (BiFPN) combined with multi-head self-attention to strengthen feature fusion across multiple resolution scales. This design enhances the model's sensitivity to small objects, such as birds and small mammals, under complex conditions including dense forests or reflective water surfaces. Meanwhile, the head of YOLOv11 utilizes a transformer-decoupled head, where classification and localization regression are performed in parallel using context-aware decoding mechanisms. This structure allows YOLOv11 to maintain high accuracy even in scenarios involving occlusion (overlapping objects) and simultaneous multi-category detection **(Alkhammash, 2025)(Sharma et al., 2024)**. Figure 4 provides a visual representation of the complete YOLOv11 architecture.

### 2.4 Model Evaluation

The evaluation of object detection model performance is a fundamental aspect of this study, as it determines how effectively the model can detect and recognize wildlife in natural environments with varying levels of visual complexity. To obtain a comprehensive assessment, a combination of quantitative metrics and qualitative visual evaluation is employed. All evaluation metrics follow standard object detection benchmarks such as the COCO Evaluation Protocol and PASCAL VOC Metrics, which are widely used in research involving YOLO, Faster R-CNN, and EfficientDet–based detection models. This study employs three primary evaluation metrics, namely precision, recall, and mean Average Precision (mAP), to assess the performance of the model.

#### 2.4.1 Precision

Precision quantifies the proportion of correct predictions (True Positives) relative to all detections produced by the model, including both correct and incorrect ones. The formula used to calculate precision is presented in Equation 1.

$$Precision = \frac{TP}{TP + FP} \quad (1)$$

where TP (True Positive) signifies the number of objects successfully and correctly detected, whereas FP (False Positive) captures the occurrences of false alarms, such as cases in which the model erroneously interprets shadows or surrounding vegetation as animal presence. A high precision value indicates that the model has strong discriminative capability, meaning it is not easily "fooled" by background objects. In the context of conservation, this is crucial to avoid false alarms in automated monitoring systems deployed in natural habitats.

#### 2.4.2 Recall

The recall metric reflects how well the model is able to detect every true target object existing in a given image. The formulation of the recall evaluation metric is shown in Equation 2.

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

where TP (True Positive) is the number of correctly detected objects, and FN (False Negative) is the number of objects that should have been detected but were missed by the model. A high recall value indicates that the model has strong sensitivity to object presence, even under complex conditions such as camouflage or low lighting.

### 2.4.3 mean Average Precision (mAP)

This research primarily relies on mAP as the central evaluation criterion, given its ability to provide a comprehensive and well-rounded measure of a model's detection and classification performance across multiple object categories. The Average Precision (AP) for each individual class is obtained by calculating the area beneath the Precision–Recall (PR) curve, as expressed in Equation 3.

$$AP = \sum (R_n - R_{n-1}) P_n \tag{3}$$

where $R_n$ and $P_n$ refer to the respective recall and precision values at each n-th threshold point. AP reflects the model's capacity to maintain elevated precision levels alongside increasing recall values. Additionally, the mathematical formulation of mAP utilized for performance evaluation in this study is provided in Equation 4.

$$mAP = \frac{1}{N} \sum_{i=1}^{N} AP_i \tag{4}$$

mAP reflects the averaged AP values obtained from all object classes present in the dataset, where N indicates the number of wildlife categories being assessed. This study employs two distinct mAP variants to comprehensively evaluate model performance. The first is mAP@0.5, which evaluates detection results at an IoU threshold of ≥ 0.5 and determines whether the spatial accuracy of predicted bounding boxes meets an acceptable standard. The second is mAP@0.5:0.95, which averages mAP scores over multiple IoU thresholds between 0.5 and 0.95 at 0.05 step intervals, providing a considerably more demanding and complete measure of spatial detection precision. The mAP metric is designated as the primary evaluation criterion to identify which between the two models, YOLOv8 or YOLOv11, achieves higher overall performance in wildlife detection within natural environments.

## 3. RESULTS AND DISCUSSION

This study uses 14 wildlife species classes collected from various natural habitats to train and evaluate the performance of two object detection models, namely YOLOv8 and YOLOv11. The dataset used in this study consists of 6,165 images after the application of data augmentation. The distribution of images across classes is initially imbalanced, which can affect model learning performance, particularly for classes with limited samples. To address this issue, data augmentation was applied proportionally across classes to increase the number of samples and improve data balance. This augmentation process enhances the diversity of visual features and improves the model's robustness against real-world environmental challenges such as occlusion, illumination variation, and complex backgrounds commonly found in natural habitats.

A comparative overview of precision and recall results is provided in Table 2 between the YOLOv8 and YOLOv11 models, while Table 3 shows the comparison results of mAP@0.5 and mAP@0.5:0.95 between the YOLOv8 and YOLOv11 models. The evaluation results indicate that both models achieve very high performance with relatively small differences between them. The quantitative evaluation reveals that YOLOv8 attains a precision of 0.997, recall of 0.982, mAP@0.5 of 0.993, and mAP@0.5:0.95 of 0.874, whereas YOLOv11 yields precision of 0.969, recall of 0.980, mAP@0.5 of 0.994, and mAP@0.5:0.95 of 0.894. These values demonstrate that both models exhibit very high accuracy, with only marginal performance differences. Class-wise performance analysis indicates that large species such as Panthera pardus melas, Sus scrofa, and Muntiacus muntjak are detected with near-perfect accuracy by both models, achieving precision and recall values close to 1.0. These species are generally easier to recognize due to their larger size and more distinctive visual features, which remain clearly visible even in complex natural environments. Although data augmentation has significantly increased the number of samples and improved overall class balance, these classes remain more challenging due to their smaller object size, partial occlusion, and visual similarity with the surrounding environment. These results suggest that while the augmentation strategy improves model generalization, limitations in visual distinctiveness and environmental complexity continue to affect detection performance for certain species.

**Table 2. Performance Comparison of YOLOv8 and YOLOv11 Based on Precision and Recall**

| Class | Precision | | Recall | |
|---|---|---|---|---|
| | YOLOv8 | YOLOv11 | YOLOv8 | YOLOv11 |
| **All** | **0.997** | **0.969** | **0.982** | **0.980** |
| Hystrix Javanica | 0.987 | 0.979 | 1 | 1 |
| Lariscus Insignis | 0.963 | 0.942 | 1 | 1 |
| Macaca Fascicularis | 0.963 | 0.982 | 0.966 | 0.942 |
| Martes Flavigula | 0.913 | 0.881 | 1 | 1 |
| Melogale Orientalis | 0.966 | 0.964 | 1 | 1 |
| Muntiacus Muntjak | 0.990 | 0.984 | 1 | 1 |
| Panthera Pardus Melas | 1 | 1 | 0.976 | 0.977 |
| Paradoxurus Hermaphoditus | 1 | 0.98 | 0.98 | 0.985 |
| Presbytis Comata | 0.958 | 0.929 | 1 | 1 |
| Prionailurus Javanensis | 1 | 1 | 0.917 | 0.975 |
| Sus Scrofa | 0.99 | 0.989 | 0.943 | 0.932 |
| Trachypithecus Auratus | 1 | 1 | 1 | 1 |
| Urva Javanica | 0.955 | 0.938 | 1 | 1 |
| Viverricula Indica | 1 | 1 | 0.967 | 0.914 |

**Table 3. Performance Comparison of YOLOv8 and YOLOv11 Based on mAP@0.5 and mAP@0.5:0.95**

| Class | mAP@0.5 | | mAP@0.5:0.95 | |
|---|---|---|---|---|
| | YOLOv8 | YOLOv11 | YOLOv8 | YOLOv11 |
| **All** | **0.993** | **0.994** | **0.874** | **0.894** |
| Hystrix Javanica | 0.995 | 0.995 | 0.995 | 0.971 |
| Lariscus Insignis | 0.995 | 0.995 | 0.970 | 0.969 |
| Macaca Fascicularis | 0.973 | 0.989 | 0.874 | 0.872 |
| Martes Flavigula | 0.995 | 0.995 | 0.995 | 0.995 |
| Melogale Orientalis | 0.995 | 0.995 | 0.651 | 0.806 |
| Muntiacus Muntjak | 0.995 | 0.995 | 0.924 | 0.903 |
| Panthera Pardus Melas | 0.995 | 0.995 | 0.931 | 0.920 |
| Paradoxurus Hermaphoditus | 0.987 | 0.994 | 0.853 | 0.856 |
| Presbytis Comata | 0.995 | 0.995 | 0.995 | 0.923 |
| Prionailurus Javanensis | 0.995 | 0.995 | 0.793 | 0.796 |
| Sus Scrofa | 0.994 | 0.986 | 0.874 | 0.847 |
| Trachypithecus Auratus | 0.995 | 0.995 | 0.964 | 0.958 |
| Urva Javanica | 0.995 | 0.995 | 0.497 | 0.796 |
| Viverricula Indica | 0.995 | 0.993 | 0.914 | 0.902 |

The small performance differences between the two models can be explained from an architectural perspective. YOLOv8 features a lightweight and efficient structure enabled by the

C2f modules and a decoupled head, which accelerates inference without sacrificing accuracy. Consequently, YOLOv8 proves to be an ideal candidate for practical real-time object detection applications conducted in natural environments through the use of camera trap devices. Meanwhile, YOLOv11 introduces improvements in feature aggregation and inter-layer connectivity, resulting in more consistent detection of small and complex objects across diverse habitat conditions. The higher mAP@0.5:0.95 achieved by YOLOv11 indicates improved spatial precision and stronger generalization capability across variations in object size.

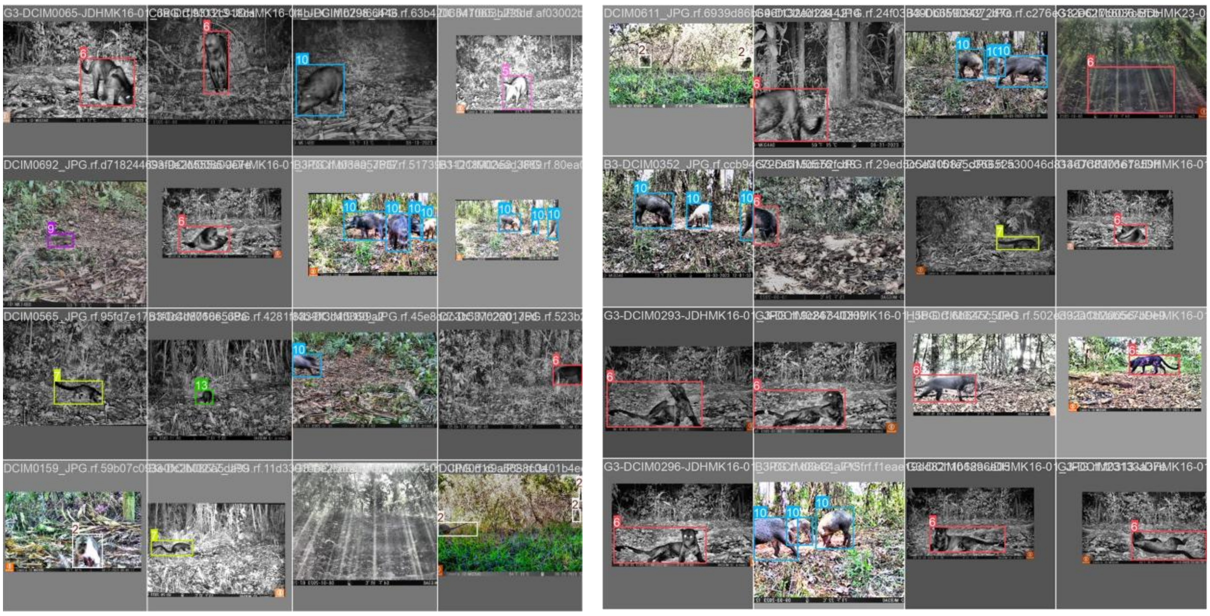

**Figure 5. Visual Comparison of YOLOv8 (left) and YOLOv11 (right) Detection Results in Natural Habitat Conditions**

To provide qualitative insight into the detection performance, Figure 5 presents a visual comparison between YOLOv8 and YOLOv11 outputs under challenging natural habitat conditions. These scenarios include small object detection, occlusion, and complex background environments with dense vegetation. As shown in the figure, YOLOv11 produces more accurate and consistent bounding box localization compared to YOLOv8. In several cases, YOLOv8 either generates less precise bounding boxes or misses smaller objects that are partially occluded or blend with the surrounding environment. In contrast, YOLOv11 is able to better capture object boundaries and maintain detection consistency even under visually complex conditions. These qualitative observations are consistent with the quantitative results, where YOLOv11 achieves a higher mAP@0.5:0.95 score, indicating superior spatial accuracy. This confirms that YOLOv11 is more effective in handling the inherent visual challenges present in wildlife detection tasks.

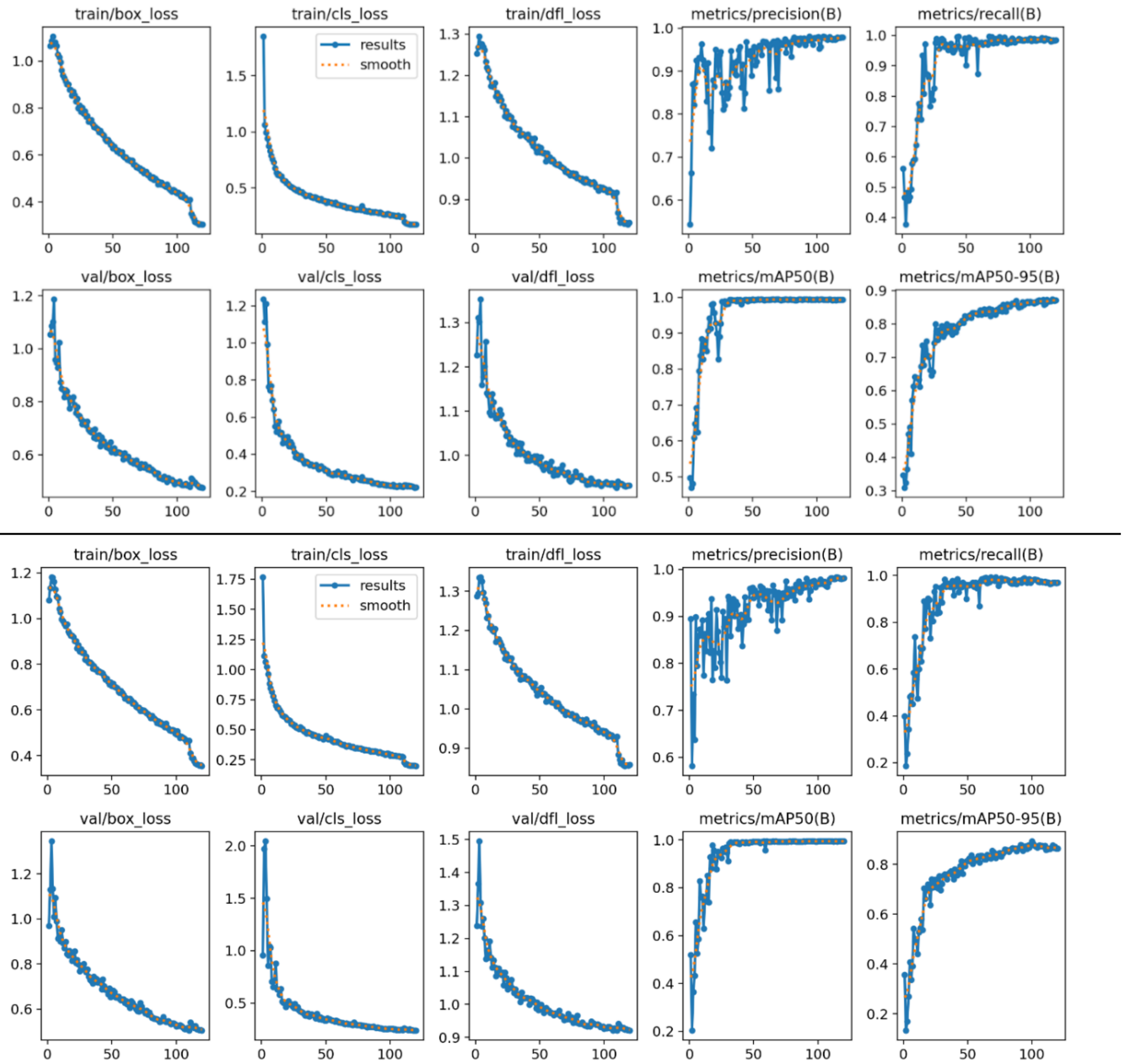


**Figure 6. Evaluation Graph – YOLOv8 (top) and YOLOv11 (bottom)**

The training and evaluation curves of the YOLOv8 model (shown in the top row) and the YOLOv11 model (shown in the bottom row) are presented in Figure 6, demonstrating consistent convergence across over 100 epochs. For both models, the training losses (box_loss, cls_loss, and dfl_loss) as well as the validation losses decrease consistently without significant divergence, indicating the absence of overfitting and good generalization capability on the validation data. YOLOv8 exhibits faster convergence with a more rapid reduction in loss during the early training phase, whereas YOLOv11 shows a smoother and more stable loss reduction pattern during the middle to later epochs. The evaluation metrics, precision and recall, increase significantly for both models and approach values close to 1.0, with YOLOv8 reaching high values earlier but with minor fluctuations, while YOLOv11 produces more stable curves. In terms of mAP@0.5, both models achieve near-perfect performance before the 50th epoch; however, for mAP@0.5:0.95, YOLOv11 attains a slightly higher final value than YOLOv8, indicating better spatial accuracy in object boundary localization.

In the context of wildlife conservation, both models demonstrate strong detection performance. However, YOLOv11 consistently provides higher spatial accuracy and is therefore considered the more effective model for detecting wildlife in complex natural environments. The achieved mAP@0.5:0.95 values above 0.89 indicate that YOLOv11 models are capable of accurately detecting objects despite challenges such as shadows, dense vegetation, and uneven lighting conditions. The application of data augmentation contributes to improved robustness by increasing data diversity and enhancing the model's ability to generalize across varying environmental conditions. This enables both models to maintain stable performance when applied to real-world camera trap data. From an application perspective, these models are suitable for automated conservation systems, supporting tasks such as wildlife population monitoring, behavioral analysis, and early detection of environmental disturbances without direct human intervention. Nevertheless, YOLOv11 demonstrates superior spatial capability, particularly in detecting small and partially occluded objects, making it more reliable for deployment in complex ecological scenarios.

## 4. CONCLUSION

A comparative investigation of two recent object detection models, YOLOv8 and YOLOv11, was conducted to evaluate their performance in wildlife detection within natural habitats. Both models were tested on 14 wildlife species classes representing the ecological and morphological diversity of tropical animals. The dataset consisted of 6,165 augmented images, improving data balance and evaluation robustness. The results show that YOLOv8 achieved precision, recall, mAP@0.5, and mAP@0.5:0.95 scores of 0.997, 0.982, 0.993, and 0.874, respectively, while YOLOv11 obtained 0.969, 0.980, 0.994, and 0.894. These findings indicate that YOLOv8 performs better in detection efficiency and precision, whereas YOLOv11 provides higher spatial accuracy and more consistent performance across different object scales. YOLOv8 is more suitable for real-time deployment, such as camera trap systems, while YOLOv11 is better for server-based analysis requiring accurate localization. Both models successfully detected large species such as Panthera pardus melas and Sus scrofa with near-perfect accuracy under diverse environmental conditions. Overall, YOLOv11 demonstrated superior performance, particularly in spatial accuracy and robustness across varying object scales, reflected by its higher mAP@0.5:0.95 score. This advantage enables YOLOv11 to better detect small, occluded, and visually complex wildlife objects in natural environments. Therefore, YOLOv11 can be considered a more effective model for wildlife detection in ecological environments with challenges such as camouflage, dense vegetation, and lighting variations.

## ACKNOWLEDGEMENT

This research was supported by Hilirisasi Riset Prioritas SINERGI program grant funded by Directorate General of Research and Development, Ministry of Higher Education, Science, and Technology (no. 70/PKS/0.0/2025) and was supported by Telkom University and Wanadri as the managing organization of Taman Buru Gunung Masigit Kareumbi.

## REFERENCES

Adhikari, B., et al. (2026). A comprehensive evaluation of YOLO-based deer detection performance on edge devices. Electronics, *15*(5), 1026. https://doi.org/10.3390/electronics15051026.

Alkhammash, E. H. (2025). A comparative analysis of YOLOv9, YOLOv10, YOLOv11 for smoke and fire detection. Fire, *8*(1), 26. https://doi.org/10.3390/fire8010026.

Charanya, P., Sridharan, S., Shankar, S. Y., Sriram, M., & Yashvanth, K. (2025). Image recognition for wildlife conservation. In 2025 International Conference on Computational, Communication and Information Technology (ICCCIT) (pp. 19–24). IEEE. https://doi.org/10.1109/ICCCIT62592.2025.10927859.

Chappidi, J., & Sundaram, D. M. (2024). Novel animal detection system: Cascaded YOLOv8 with adaptive preprocessing and feature extraction. IEEE Access, *12*, 110575–110587. https://doi.org/10.1109/ACCESS.2024.3439230.

Chen, L., Li, G., Zhang, S., Mao, W., & Zhang, M. (2024). YOLO-SAG: An improved wildlife object detection algorithm based on YOLOv8n. Ecological Informatics, *83*, 102791. https://doi.org/10.1016/j.ecoinf.2024.102791.

Corputty, F., Wibowo, S. A., & Rizal, S. (2022). Implementation of object detection and recognition based on exploration deep neural network features for quadcopter. In Proceedings of the International Conference on Information and Communications Technology (ICOIACT) (pp. 485–490). IEEE. https://doi.org/10.1109/ICOIACT55506.2022.9971943.

Corputty, F., Wibowo, S. A., Sunarya, U., Rahmania, R., & Hidayat, S. W. (2024). Enhancing retail product recognition using modified YOLOv8 and self-supervised learning. Journal of Theoretical and Applied Information Technology, *102*(13), 5158–5167.

Fang, S., Shen, Y., Zou, H., Yin, Y., Jin, W., & Zhou, H. (2025). Birds-YOLO: A bird detection model for Dongting Lake based on modified YOLOv11. Biology, *14*(11), 1515. https://doi.org/10.3390/biology14111515.

Feng, F., Hu, Y., Li, W., & Yang, F. (2024). Improved YOLOv8 algorithms for small object detection in aerial imagery. Journal of King Saud University – Computer and Information Sciences, *36*(6), 102113. https://doi.org/10.1016/j.jksuci.2024.102113.

Gokulapriya, S., & Princy Suganthi, B. S. (2025). Deep learning framework for real-time animal detection in rural surveillance applications. In Proceedings of the International Conference on Creative Communication and Innovative Technology (ICCIT) (pp. 1–6). IEEE. https://doi.org/10.1109/ICCIT65724.2025.11167423.

Jiang, H., Zhao, J., Ma, F., Yang, Y., & Yi, R. (2025). Mobile-YOLO: A lightweight object detection algorithm for four categories of aquatic organisms. Fishes, *10*(7), 348. https://doi.org/10.3390/fishes10070348.

Jiang, L., & Wu, L. (2024). Enhanced YOLOv8 network with extended Kalman filter for wildlife detection and tracking in complex environments. Ecological Informatics, *84*, 102856. https://doi.org/10.1016/j.ecoinf.2024.102856.

Jiang, P., Ergu, D., Liu, F., Cai, Y., & Ma, B. (2022). A review of YOLO algorithm developments. Procedia Computer Science, *199*, 1066–1073. https://doi.org/10.1016/j.procs.2022.01.135.

John, R., & Issac, J. J. (2025). Wildlife monitoring using YOLOv8 and edge-AI: A real-time approach. IUP Journal of Telecommunications, *17*(3), 20–32. https://doi.org/10.71329/IUPJTC/2025.17.3.20-32

Karanam, V. K., Kommuri, V. S. A., Lekkala, H. V. R., & Dhanraj, J. A. (2026). Bird species recognition using YOLOv8: A deep learning approach for habitat conservation and preservation. Computer Vision and Robotics, *1644*, 331–341. https://doi.org/10.1007/978-3-032-06253-6_26.

Lawal, O. M., Tan, Y., & Liu, C. (2025). MADNet: Marine animal detection network using the YOLO platform. PLOS ONE, *20*(5), e0322799. https://doi.org/10.1371/journal.pone.0322799.

Li, J., et al. (2025). YOLOv11-OPNet: Dual-modality framework for fishing vessel detection in complex long-range maritime scenarios. Aquaculture International, *33*(6), 533. https://doi.org/10.1007/s10499-025-02216-0.

Ma, D., & Yang, J. (2022). YOLO-Animal: An efficient wildlife detection network based on improved YOLOv5. In 2022 International Conference on Image Processing, Computer Vision and Machine Learning (ICICML) (pp. 464–468). IEEE. https://doi.org/10.1109/ICICML57342.2022.10009855.

May, G., Dalsasso, E., Delplanque, A., Kellenberger, B., & Tuia, D. (2025). How to minimize the annotation effort in aerial wildlife surveys. Ecological Informatics, *91*, 103387. https://doi.org/10.1016/j.ecoinf.2025.103387.

Mughal, A. B., Khan, R. U., Rehman, A. U., & Bermak, A. (2025). Deep learning for dynamic wildlife monitoring: A real-time approach. IEEE Access, *13*, 147422–147448. https://doi.org/10.1109/ACCESS.2025.3600625.

Naidu, A. P., Gosalia, H., Gakhar, I., Rathore, S. S., Didwania, K., & Verma, U. (2025). DEAL-YOLO: Drone-based efficient animal localization using YOLO. arXiv. https://arxiv.org/abs/2503.04698

O'Reilly, V., Cox, J., Leontidis, G., Hanheide, M., Bosilj, P., & Brown, J. (2025). MMDEW: Multipurpose multiclass density estimation in the wild. arXiv. https://doi.org/10.48550/arXiv.2503.04698.

Prabu, V. C., Vidya, Q. M. M., Sathiyamoorthi, V., Durgadevi, P., & Gowthami, M. (2026). A novel object detection system for computer night vision images using residual 3D transformer-based YOLOv8 with adaptive GRU in edge and cloud sector. Expert Systems with Applications, *298*, 129554. https://doi.org/10.1016/j.eswa.2025.129554.

Roca, A., et al. (2025). Detection of endangered deer species using UAV imagery: A comparative study between efficient deep learning approaches. In Proceedings of the International Conference on Unmanned Aircraft Systems (ICUAS), (pp. 83–90). IEEE. https://doi.org/10.1109/ICUAS65942.2025.11007886.

Roy, A. M., Bhaduri, J., Kumar, T., & Raj, K. (2023). WilDect-YOLO: An efficient and robust computer vision-based accurate object localization model for automated endangered wildlife detection. Ecological Informatics, *75*, 101919. https://doi.org/10.1016/j.ecoinf.2022.101919.

Sharma, A., Kumar, V., & Longchamps, L. (2024). Comparative performance of YOLOv8, YOLOv9, YOLOv10, YOLOv11 and Faster R-CNN models for detection of multiple weed species. Smart Agricultural Technology, *9*, 100648. https://doi.org/10.1016/j.atech.2024.100648

Yang, W., et al. (2023). A forest wildlife detection algorithm based on improved YOLOv5s. Animals, *13*(19), 3134. https://doi.org/10.3390/ani13193134.

Zhang, Y., & Cai, Z. (2023). CE-RetinaNet: A channel enhancement method for infrared wildlife detection in UAV images. IEEE Transactions on Geoscience and Remote Sensing, *61*, 1–12. https://doi.org/10.1109/TGRS.2023.3299651.

Zhao, H., Jia, L., Wang, Y., & Yan, F. (2025). Autonomous UAV detection of Ochotona curzoniae burrows with enhanced YOLOv11. Drones, *9*(5), 340. https://doi.org/10.3390/drones9050340.

Zhong, J., Li, M., Qin, J., Cui, Y., Yang, K., & Zhang, H. (2022). Real-time marine animal detection using YOLO-based deep learning networks in the coral reef ecosystem. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, *41*(3), 301–306. https://doi.org/10.5194/isprs-archives-XLVI-3-W1-2022-301-2022.

Zhu, Y., Zhao, Y., He, Y., Wu, B., & Su, X. (2025). YOLO-WildASM: An object detection algorithm for protected wildlife. Animals, *15*(18), 2699. https://doi.org/10.3390/ani15182699.