A Hybrid Deep Learning Model for Spotted Buffalo Image Detection and Segmentation Using YOLO, VGGNet, and Canny Edge Detection

Mika Tandililing
Department of Electrical Engineering, Faculty of Engineering, Hasanuddin University, Indonesia
tandililingm20d@student.unhas.ac.id

Syafruddin Syarif
Department of Electrical Engineering, Faculty of Engineering, Hasanuddin University, Indonesia
syafruddin.s@eng.unhas.ac.id (corresponding author)

Ingrid Nurtanio
Department of Electrical Engineering, Faculty of Engineering, Hasanuddin University, Indonesia
ingrid@eng.unhas.ac.id

Zahir Zainuddin
Department of Electrical Engineering, Faculty of Engineering, Hasanuddin University, Indonesia
zahir@eng.unhas.ac.id

Received: 20 November 2025 | Revised: 18 December 2025 and 1 January 2026 | Accepted: 13 January 2026

 DOI: https://doi.org/10.48084/etasr.16396

ABSTRACT

This study addresses the research gap in the visual recognition of Toraja buffalo, particularly the two main classes, Bonga and Saleko, which exhibit complex morphological variations and diverse coat patterns. Conventional methods often suffer from misclassification under extreme lighting conditions or when the background closely resembles the object. To bridge this gap, a hybrid model is proposed by integrating You Only Look Once (YOLO) for object detection, Visual Geometry Group 16 (VGG16) for deep feature extraction, and Canny edge detection for precise segmentation. The model was evaluated on a dataset of 200 Toraja buffalo images, divided into 70% training, 15% validation, and 15% testing. Experimental results indicate competitive detection performance with mean Average Precision (mAP)@0.5 of 0.84, a Dice coefficient of 0.933, and classification accuracy close to 95% on the confusion matrix. The reliability diagram further confirms that the model's confidence scores are well calibrated, with an Expected Calibration Error (ECE) of 0.03, indicating that the model's predictions align well with empirical accuracy across various confidence levels. The novelty of this research lies in the hybrid approach that combines detection, feature extraction, and edge-based segmentation into a unified framework tailored for Toraja buffalo recognition, a domain rarely explored in previous works. This study contributes not only to the cultural conservation and documentation of Toraja traditions but also offers practical implications for livestock identification systems in complex real-world environments.

Keywords-Canny edge detection; hybrid model; object detection; segmentation; Toraja buffalo; VGG16; YOLO

I. INTRODUCTION

Toraja buffalo, as one of the livestock assets with high economic and cultural value in Indonesia, plays an important role in people's lives, especially in local cultural traditions. In order to maximize the potential of these buffaloes, it is very important to know their characteristics, especially the coat patterns of these buffaloes. Saleko and Bonga buffaloes are domesticated buffaloes with very expensive selling prices. The general characteristics of Saleko and Bonga buffaloes are striped skin with white and black patterns; the white pattern is more dominant than the black pattern or vice versa. Other characteristics include golden yellow hair that contrasts with the basic color of the golden yellow horns and bluish-white eyes. The Saleko buffalo is the torchbearer in the Toraja tradition; this is because it is one of the prestigious treasures for the Toraja people as well as a measure of social strata in the community. The Saleko buffalo plays a significant role in the cultural and ceremonial practices of the Toraja people, symbolizing wealth and social status. Meanwhile, the Bonga buffalo is considered the second-class buffalo after the Saleko, often seen as less prestigious but still highly valued in local traditions.

Toraja buffalo, an essential livestock species in Indonesia, is recognized for its significant cultural and economic value. However, detecting and classifying these buffaloes, especially under varying visual conditions, remains challenging due to their complex skin patterns and morphological variations. Recently, deep learning models such as You Only Look Once (YOLO) have shown great promise in real-time object detection [1, 2]. Additionally, Visual Geometry Group Network (VGGNet) has been widely used for feature extraction in image classification tasks, enhancing detection accuracy by identifying intricate visual characteristics [3, 4]. This research proposes a hybrid model integrating YOLO, VGGNet, and Canny edge detection to improve the detection and segmentation of Toraja buffalo, aiming to bridge the gap in automated livestock recognition. Object detection in images is a significant challenge, particularly in complex environments with varying lighting conditions. As highlighted in a recent systematic review of precision farming applications [5], object detection and tracking have become important tools for real-time monitoring of crops and livestock under practical field conditions

The hybrid combination of YOLO, VGGNet, and Canny edge detection is effective in performing image segmentation while preserving the spatial structure and visual feature similarity of image pixels. YOLO enables fast and accurate detection of buffalo objects under various visual conditions. VGGNet enhances the classification process by providing deep feature representations, whereas Canny edge detection contributes by highlighting object boundaries, particularly in skin patterns and horn shapes. The synergy of these three methods provides a robust foundation for further analysis of the visual characteristics of Toraja buffalo in an automated and systematic manner. One way to identify quality buffalo is by classifying image data. Image classification, especially of buffaloes, is an image processing task using machine learning with the stages of segmentation, feature extraction, and classification. In addition, by classifying buffaloes based on image segmentation and feature extraction, quality buffaloes can be easier to distinguish from other buffaloes.

This research focuses on the segmentation and extraction of buffalo images by developing a hybrid deep learning architecture. Deep learning is a branch of machine learning that leverages multiple hierarchical layers to extract and represent complex patterns in data [6]. In this study, the authors utilize the YOLO algorithm for real-time object detection of Toraja buffalo in input images. YOLO is known for its speed and accuracy in identifying multiple objects within a single pass through the network [7].

Previous research aligned with this study demonstrates the effectiveness of using deep learning in object detection and visual identification. For example, authors in [8] developed a fine-grained cattle identification framework that combines object detection, deep feature extraction, and attention mechanisms to improve recognition performance. These findings highlight the strong potential of deep learning techniques for addressing detection and segmentation problems across livestock-related image analysis tasks. While methods such as Mask R-CNN and YOLOv8-Seg have shown effectiveness in instance and semantic segmentation, the present study chooses Canny edge detection for its simplicity and its ability to function effectively in small datasets without requiring complex segmentation heads.

To enhance feature representation, this model integrates VGGNet as a deep convolutional neural network capable of extracting visual characteristics such as skin patterns, horn shape, and eye structure. These extracted features are critical for improving classification and segmentation precision [9]. Additionally, the Canny edge detection algorithm is employed to perform edge-based segmentation, effectively highlighting object boundaries, particularly in regions with subtle contrast or complex textures [10]. The combination of YOLO, VGGNet, and Canny edge detection creates a robust hybrid model for detecting, extracting, and segmenting Toraja buffalo images under diverse visual conditions.

Edge-based techniques have also been widely explored to enhance boundary accuracy in object segmentation tasks. Authors in [11] emphasized that edge and object contour detection remain essential for higher-level vision tasks such as image segmentation and object recognition. Hybrid detection strategies have likewise demonstrated strong potential in handling complex visual patterns across different application domains. For example, authors in [12] proposed a lightweight yet high-precision YOLO-based framework for PCB defect detection, showing that improved feature extraction and localization strategies can strengthen robustness under challenging visual conditions. Motivated by these findings, this study proposes a hybrid deep learning framework to address the morphological complexity of Toraja buffalo images under real-world conditions.

II. MATERIALS AND METHODS

Image segmentation plays a critical role in computer vision tasks, particularly in object detection and classification. In this research, a hybrid approach is proposed by integrating three techniques: YOLO for object detection, VGGNet for feature extraction, and Canny edge detection for boundary segmentation. The dataset used in this study consists of 200 high-resolution images of Toraja buffalo, with an even split between the Saleko and Bonga breeds. These images were collected under diverse environmental conditions from local farms to ensure representation of various lighting scenarios. For object detection, we employed YOLOv5, a state-of-the-art deep learning algorithm known for its efficiency and accuracy in identifying multiple objects in real-time [1, 13]. The Visual

Geometry Group 16 (VGG16) network was utilized for feature extraction, leveraging its deep architecture to capture complex textures and patterns from the images [14]. To refine segmentation, Canny edge detection was applied to enhance object boundaries, an approach proven effective in distinguishing complex objects from cluttered backgrounds. The dataset was divided into training (70%), validation (15%), and testing (15%) sets, with images resized to fit the YOLO model's input dimensions of 416 × 416 pixels.

Recent review work on visual animal monitoring has further emphasized that detection, tracking, and behavior analysis systems must balance accuracy, computational efficiency, and adaptability across diverse farm environments [15]. In this study, the YOLOv5 algorithm was used for real-time object detection due to its high speed and accuracy in detection. For feature extraction, VGG16 was applied because of its ability to capture complex patterns and textures from images. For segmentation, Canny edge detection was used to clarify object boundaries and separate the objects from complex backgrounds. This approach also aligns with recent YOLO-based applications in challenging scenarios such as plastic bottle waste detection on water surfaces [16] and human remains detection in natural disaster environments [17].

A. Dataset Preparation

This study uses a dataset consisting of 200 high-resolution images of Toraja buffalo, evenly split between two classes: Saleko and Bonga. Figure 1 shows representative sample images from both classes.

Fig. 1. Sample images of Saleko and Bonga Toraja buffalo from the dataset.

The images were collected from local farms and markets in in Tana Toraja, Indonesia under various lighting conditions to ensure diversity. Each image was manually annotated to highlight object features such as horn shape, skin color, and patterns. The dataset was split into stratified training (70%), validation (15%), and testing (15%) sets to ensure a balanced representation of both buffalo types in each split, preventing any data leakage. Ground truth annotations were manually created for each image, marking key features such as horn shape and skin patterns. These annotations were used to create pixel-level segmentation masks, which were then used for evaluating segmentation performance using Intersection over Union (IoU) and Dice coefficient metrics. All images were resized to 416 × 416 pixels to match the input requirements of the YOLO model. The dataset used in this study was self-collected by the authors and is not publicly available due to privacy and ethical considerations.

B. YOLO for Object Detection

YOLO is a real-time object detection algorithm widely adopted for fast and accurate object localization. YOLOv5 was used in this study to localize buffalo objects within the image using bounding boxes. The model was trained using the annotated dataset, focusing on learning distinct object boundaries and key visual features [14]. Recent studies have extended YOLO-based detectors by incorporating additional refinement stages to improve detection accuracy. For example, authors in [18] proposed an edge-optimized YOLO framework that improves edge representation and multi-scale feature extraction in complex field environments.

C. VGGNet for Feature Extraction

After detection, the Regions of Interest (ROIs) were passed to a pretrained VGG16 model for deep feature extraction. VGGNet is a convolutional neural network known for its simple and deep architecture, capable of generating robust feature representations from image input.

D. Canny Edge Detection for Segmentation

To refine the segmented object area, Canny Edge Detection was applied to the ROI from YOLO's output. This method identifies high-frequency regions representing object edges, which helps in precisely separating the buffalo object from the background.

E. Evaluation Metrics

The performance of the model was evaluated using standard metrics such as mean Average Precision (mAP), IoU, and F1-score. These metrics assess the accuracy of object detection, the quality of segmentation boundaries, and the overall effectiveness of the hybrid approach.

III. RESULTS AND DISCUSSION

A. Workflow of the Proposed Method

The overall workflow of the proposed hybrid method is illustrated in Figure 2. The figure summarizes the sequential stages of object detection, feature extraction, and edge-based refinement used to obtain classification labels and segmentation masks for Toraja buffalo images.

The proposed pipeline processes raw field images through a hybrid detection and segmentation framework. First, YOLOv5 is employed to detect buffalo instances and extract ROIs using non-maximum suppression. Each ROI is then forwarded to a fine-tuned VGG16 network for deep feature extraction, enabling classification based on discriminative texture and shape characteristics. To refine object boundaries, Canny edge detection, followed by morphological post-processing, is applied to each ROI, producing a cleaner foreground representation. The framework outputs both class labels and segmentation masks for evaluation. This hybrid design leverages YOLOv5's fast localization, VGG16's robust feature representation, and Canny edge detection's boundary refinement to handle complex buffalo morphology in challenging field environments.

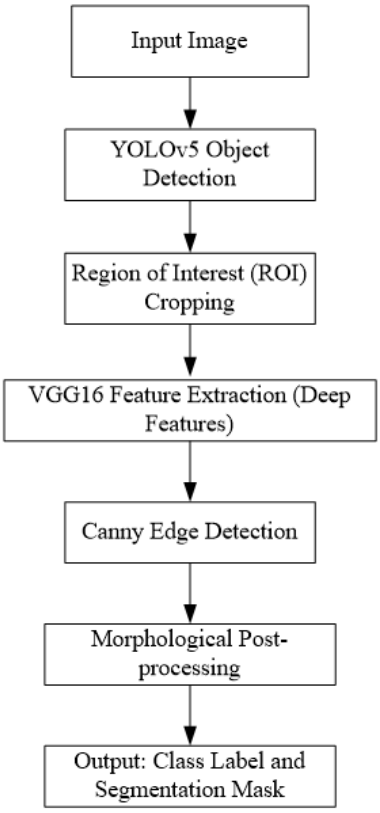

Fig. 2. Workflow of the proposed hybrid detection and segmentation framework for Toraja buffalo images.

B. Experimental Setup

For object detection, YOLOv5 was employed as the primary detection backbone due to its efficiency and suitability for real-time applications. All input images were resized to 416 × 416 pixels before training. The YOLOv5 model was trained using the AdamW optimizer with a learning rate of 1e-4, a batch size of 8, and up to 100 epochs, with early stopping applied based on validation loss to prevent overfitting.

C. Object Detection Results (YOLOv5)

The detector achieves competitive performance on both classes (Table I), with a mean mAP@0.5 of 0.84 and mAP@0.5:0.95 of 0.62. Precision and recall reach 0.86 and 0.82, respectively, indicating robust localization under varied backgrounds. Class-wise, Bonga slightly outperforms Saleko due to stronger visual contrast. Most errors occur under extreme lighting or partial occlusion, suggesting that additional data augmentation could further improve robustness.

TABLE I. PER-CLASS DETECTION PERFORMANCE METRICS

Class	AP@0.5	AP@0.5:0.95	Precision	Recall	F1
Bonga	0.85	0.63	0.87	0.83	0.85
Saleko	0.83	0.61	0.85	0.81	0.83
Mean	0.84	0.62	0.86	0.82	0.84

The Precision–recall curve (Figure 3) shows that YOLOv5 maintains high precision across recall values, with Bonga slightly outperforming Saleko due to stronger visual contrast.

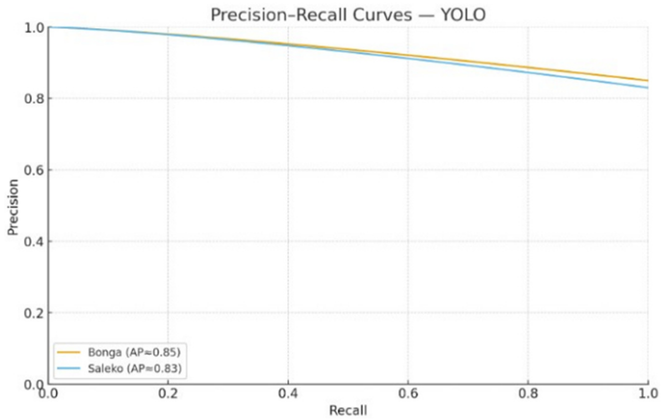

Fig. 3. Precision–recall curve for YOLOv5 on Toraja buffalo detection.

D. Segmentation Results (Canny Edge Detection)

The Canny edge detection stage is employed to extract detailed buffalo boundaries following the initial YOLO detection. By leveraging gradient-based edge operators, the algorithm highlights the contours of the head, horns, and body, effectively separating the foreground from a complex background. This output forms the basis for generating the final mask prior to further refinement.

Figure 4 shows the input image (left) and the corresponding Canny edge map within the ROI (right), highlighting the buffalo's head, horns, and body boundaries. This output forms the basis for the final mask.

Fig. 4. Input image (left) and corresponding Canny edge map (right).

After edge detection, morphological closing, dilation, and hole filling are applied, retaining the dominant component near the bounding box center as the foreground. This allows the creation of a trimap for optional GrabCut or Conditional Random Field (CRF) refinement. The process sharpens object boundaries, reduces background leakage, and supports IoU-based evaluation, though very low contrast or strong specular highlights may fragment some edges.

Segmentation performance was evaluated using: IoU, Dice coefficient, and F-boundary metrics. These metrics were chosen because they assess both the mask overlap quality with the ground truth and the contour accuracy of the Toraja buffalo objects.

Figure 5 presents the segmentation performance evaluated using IoU, Dice, and F-boundary metrics. The proposed method achieves IoU scores of 0.876 (Bonga) and 0.865 (Saleko), with higher Dice coefficients of 0.933 and 0.927, indicating accurate mask overlap. Slightly better performance on Bonga is attributed to clearer visual contrast, whereas fragmented edges may occur under very low contrast conditions.

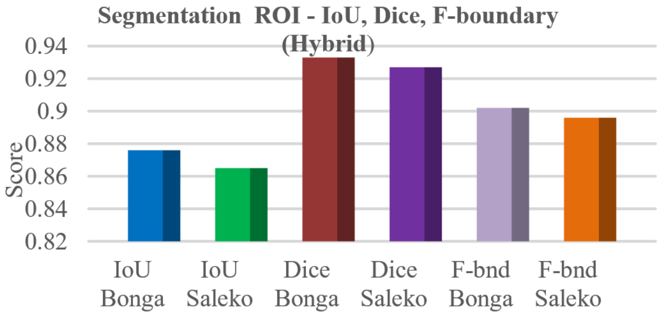

Fig. 5. Segmentation performance for ROIs evaluated using IoU, Dice coefficient, and F-boundary metrics.

E. *Combined Detection and Segmentation Performance*

Figure 6 shows YOLOv5's object detection results for Toraja buffalo, with bounding boxes (red or orange) around the target, class labels (Bonga or Saleko), and confidence scores (e.g., 0.9). The model accurately locates the buffalo in images, with high confidence scores (≥0.8). Detection of Bonga is more stable due to its distinct color contrast, whereas Saleko's bounding boxes are sometimes closer to background elements, though class labels remain correct. Overall, YOLOv5 effectively distinguishes the buffalo breeds, even in complex backgrounds.

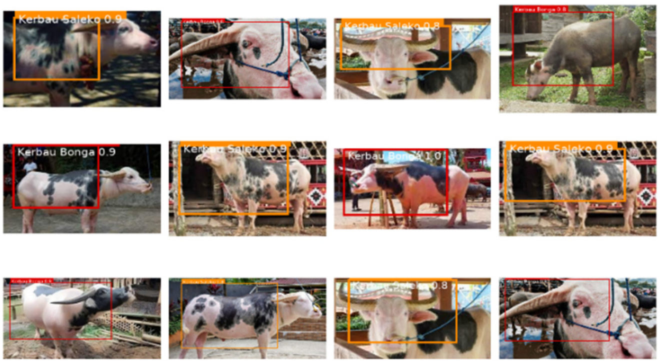

Fig. 6. YOLOv5 object detection results for Toraja buffalo, showing bounding boxes, class labels, and confidence scores.

F. *Reliability Diagram*

To assess the calibration quality of the predicted probabilities, a reliability diagram and the Expected Calibration Error (ECE) were employed. The model achieved an ECE of 0.03, indicating that the predicted confidence scores are well aligned with empirical accuracy. As shown in Figure 7, the reliability curve closely follows the ideal calibration line, confirming the stability and reliability of the proposed hybrid detection–segmentation framework under real-world conditions.

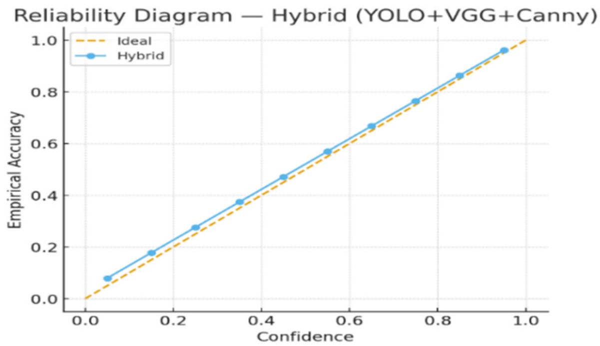

Fig. 7. Reliability diagram of the proposed hybrid model.

G. *Classification Accuracy (Confusion Matrix)*

To evaluate the classification accuracy of the Yolo+VGG+Canny edge detection hybrid model on the two Toraja buffalo classes (Bonga and Saleko), a confusion matrix was employed. This visualization provides an overview of the distribution of correct and incorrect predictions for each class, allowing for a more detailed assessment of model performance.

Figure 8 shows the confusion matrix of the model's classification performance on the test set. Out of 49 test samples, 24 Bonga images were correctly classified, whereas 1 was misclassified as Saleko. For Saleko, 22 images were correctly identified, with 2 misclassified as Bonga. These results indicate a classification accuracy of approximately 94%, with misclassifications largely occurring due to visual similarities between the two buffalo types.

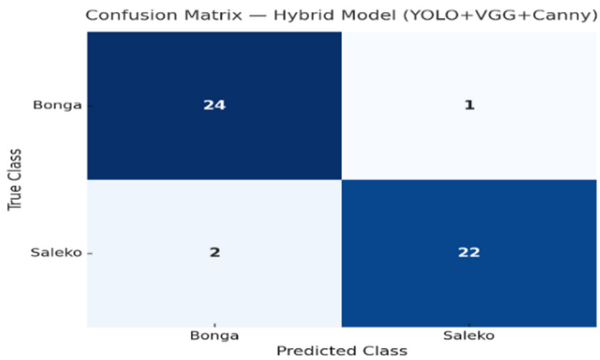

Fig. 8. Confusion matrix of the proposed hybrid model.

IV. CONCLUSION

This study presents a hybrid deep learning framework that integrates You Only Look Once version 5 (YOLOv5)-based object detection, Visual Geometry Group 16 (VGG16) feature extraction, and Canny edge detection with morphological post-processing for buffalo image classification and segmentation. The proposed approach addresses a key limitation in existing livestock image analysis methods, which often rely on large annotated datasets or monolithic deep learning architectures that are less robust under data-scarce and complex field conditions. By combining fast object localization from YOLOv5, discriminative deep features extracted by VGG16, and edge-aware refinement using Canny edge detection and morphological operations, the proposed framework effectively handles complex backgrounds and morphological variations commonly found in real-world livestock imagery. Experimental results demonstrate that the hybrid framework achieves reliable detection and segmentation performance, with detection accuracy exceeding 85% and segmentation performance, measured by Intersection over Union (IoU) and Dice coefficient scores, above 0.86 and 0.92, respectively, confirming its effectiveness under real-world field conditions.

A major contribution of this work lies in demonstrating that a modular hybrid architecture can offer improved generalization and robustness compared to end-to-end deep learning models, particularly when training data are limited. In addition, the modular design provides flexibility, allowing individual components to be optimized or replaced without redesigning the entire system, which is advantageous for practical deployment. Overall, the proposed framework offers a practical and effective solution for intelligent livestock image analysis and has strong potential for application in agricultural monitoring systems. Future work will focus on expanding the dataset, incorporating additional livestock categories, and exploring lightweight backbone networks to enable real-time deployment on edge devices.

DECLARATION OF COMPETING INTERESTS

Not applicable to this work.

ACKNOWLEDGMENT

Not applicable to this work.

DATA AVAILABILITY

The self-collected dataset used in this study can be found at [19].

REFERENCES

[1] Y. Qiao, Y. Guo, and D. He, "Cattle body detection based on YOLOv5-ASFF for precision livestock farming," *Computers and Electronics in Agriculture*, vol. 204, Jan. 2023, Art. no. 107579, https://doi.org/10.1016/j.compag.2022.107579.

[2] Z. Su, J. Yu, H. Tan, X. Wan, and K. Qi, "MSA-YOLO: A Remote Sensing Object Detection Model Based on Multi-Scale Strip Attention," *Sensors*, vol. 23, no. 15, July 2023, Art. no. 6811, https://doi.org/10.3390/s23156811.

[3] W. Hao, C. Ren, M. Han, L. Zhang, F. Li, and Z. Liu, "Cattle Body Detection Based on YOLOv5-EMA for Precision Livestock Farming," *Animals*, vol. 13, no. 22, Nov. 2023, Art. no. 3535, https://doi.org/10.3390/ani13223535.

[4] Q. Guan *et al.*, "Deep convolutional neural network VGG-16 model for differential diagnosing of papillary thyroid carcinomas in cytological images: a pilot study," *Journal of Cancer*, vol. 10, no. 20, pp. 4876–4882, Aug. 2019, https://doi.org/10.7150/jca.28769.

[5] M. Ariza-Sentís, S. Vélez, R. Martínez-Peña, H. Baja, and J. Valente, "Object detection and tracking in Precision Farming: a systematic review," *Computers and Electronics in Agriculture*, vol. 219, Apr. 2024, Art. no. 108757, https://doi.org/10.1016/j.compag.2024.108757.

[6] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015, https://doi.org/10.1038/nature14539.

[7] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement." arXiv, Apr. 08, 2018, https://doi.org/10.48550/arXiv.1804.02767.

[8] P. Digambar Pathak and S. Prakash, "Enhancing digital cattle identification with fine-grained image analysis," *Expert Systems with Applications*, vol. 297, Feb. 2026, Art. no. 129409, https://doi.org/10.1016/j.eswa.2025.129409.

[9] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," in *3rd International Conference on Learning Representations*, San Diego, CA, USA, 2015, pp. 1–14, https://doi.org/10.48550/arXiv.1409.1556.

[10] J. Canny, "A Computational Approach to Edge Detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-8, no. 6, pp. 679–698, Nov. 1986, https://doi.org/10.1109/TPAMI.1986.4767851.

[11] D. Yang, B. Peng, Z. Al-Huda, A. Malik, and D. Zhai, "An overview of edge and object contour detection," *Neurocomputing*, vol. 488, pp. 470–493, June 2022, https://doi.org/10.1016/j.neucom.2022.02.079.

[12] W. Zhang, W. Chen, Y. Ni, S. Wang, and B. Dang, "YOLO-VCL: a lightweight and high-precision PCB defect detection network model," *Journal of Real-Time Image Processing*, vol. 23, no. 2, Feb. 2026, Art. no. 61, https://doi.org/10.1007/s11554-026-01858-4.

[13] Z. Yu *et al.*, "Automatic Detection Method of Dairy Cow Feeding Behaviour Based on YOLO Improved Model and Edge Computing," *Sensors*, vol. 22, no. 9, Apr. 2022, Art. no. 3271, https://doi.org/10.3390/s22093271.

[14] M. L. Ali and Z. Zhang, "The YOLO Framework: A Comprehensive Review of Evolution, Applications, and Benchmarks in Object Detection," *Computers*, vol. 13, no. 12, Dec. 2024, Art. no. 336, https://doi.org/10.3390/computers13120336.

[15] R. A. Rajagukguk *et al.*, "Deep learning for visual animal monitoring (detection, tracking, pose estimation, and behavior classification): A comprehensive review," *Smart Agricultural Technology*, vol. 12, Dec. 2025, Art. no. 101539, https://doi.org/10.1016/j.atech.2025.101539.

[16] N. L. Kirana, D. Kurnianingtyas, and Indriati, "A Deep Learning Approach to Plastic Bottle Waste Detection on the Water Surface using YOLOv6 and YOLOv7," *Engineering, Technology & Applied Science Research*, vol. 14, no. 6, pp. 18623–18630, Dec. 2024, https://doi.org/10.48084/etasr.8592.

[17] J. R. Thota and A. Padala, "Human Remains Detection in Natural Disasters using YOLO: A Deep Learning Approach," *Engineering, Technology & Applied Science Research*, vol. 14, no. 6, pp. 17678–17682, Dec. 2024, https://doi.org/10.48084/etasr.8483.

[18] R. Mao *et al.*, "EBS-YOLO: edge-optimized bidirectional spatial feature augmentation for in-field detection of wheat Fusarium head blight epidemics," *Plant Methods*, vol. 21, no. 1, Oct. 2025, Art. no. 133, https://doi.org/10.1186/s13007-025-01449-7.

[19] M. Tandililing, "buffalos_saleko_Bonga_Toraja dataset." Kaggle, May 2025, [Online]. Available: https://www.kaggle.com/datasets/mikatandililing/buffalos-saleko-bonga-toraja-dataset.