YOLOv8-Based Deep Learning for Multi-Class Farmland Intrusion Detection: Architecture, Transfer Learning, and Wavelet Coefficient Shrinkage Preprocessing

[1*]Stella Kehinde Ogunkan, [2]Olusegun Olajide Adeosun, [3]Stephen Olatunde Olabiyisi, [4]Rantiola Fidelis Famutimi, [5]Ojo Stephen Aderibigbe

[1,2,3]Department of Computer Science, Ladoke Akintola University of Technology, Ogbomoso, Nigeria

[4]Department of Mathematics and Computer Sciences, University of Medical Sciences, Ondo, Nigeria

[5]Department of Computer Sciences, Lagos State University of Science and Technology, Ikorodu, Lagos State, Nigeria

DOI: https://doi.org/10.47772/IJRISS.2026.100600227

Received: 29 May 2026; Accepted: 03 June 2026; Published: 20 June 2026

ABSTRACT

The computational vision problem of object detection and classification in outdoor agricultural settings is particularly difficult due to the combination of highly variable illumination, cluttered backgrounds, multi-scale target sizes, class imbalance, and sensor noise from low-cost IoT camera nodes. In this paper, a YOLOv8-based Convolutional Neural Network (CNN) pipeline for real-time multi-class agricultural incursion detection is designed, implemented, and rigorously evaluated. Images are categorized by the system into three security-related groups: No-Intrusion, Animal (cows, goats, and livestock), and Human Intruder. While domain-specific fine-tuning on a bespoke 1,850-image field dataset adjusts the model to agricultural circumstances, transfer learning from the Microsoft COCO dataset (330,000+ annotated images, 80 object categories) offers fundamental feature representations. Daubechies-1 (db1) wavelets with two-level decomposition are used in a specific wavelet-based picture preprocessing pipeline and soft coefficient thresholding maintains important object borders and textures while reducing sensor and ambient noise. The system obtained a mAP@50 of 98.58% on the held-out test set after APO-based hyperparameter optimization, with per-class precision above 97% in all three categories. While the Animal class showed a recall of 94.51%, which represents a 29.11 percentage point gain over the 65.4% baseline recall and directly addresses the most important safety gap in previous farmland security systems, the Human Intruder class attained perfect memory (100%). Confusion matrix study confirms operational suitability for production IoT deployment with a remarkably low false positive rate of 0.72%.

Keywords: YOLOv8, Object Detection, Transfer Learning, Wavelet Coefficient Shrinkage, Daubechies Wavelet

INTRODUCTION

With superhuman performance on the ImageNet benchmark (He et al., 2016) and exceptional adaptability in a variety of fields, including autonomous driving and medical imaging, Convolutional Neural Networks have revolutionized computer vision. Redmon et al. (2016) introduced the YOLO (You Only Look Once) family of architectures, which represent a paradigm shift in object detection. By reformulating detection as a single regression problem—predicting bounding boxes and class probabilities from full images simultaneously in one evaluation pass—YOLO architectures achieve real-time inference speeds that are unachievable by region-proposal-based methods (Faster R-CNN, Mask R-CNN) while maintaining competitive accuracy.

The most recent version from Ultralytics (2023), YOLOv8, offers significant architectural improvements over its predecessors: an anchor-free detection head removes the sensitivity to specified anchor box shapes that made YOLOv5 deployment challenging in non-standard visual domains; A decoupled head architecture

divides the classification and regression branches for better training dynamics, while a modified C2f (Cross-Stage Partial Network with two bottleneck blocks) backbone enhances gradient flow and feature reuse. Because of these advancements, YOLOv8 is especially well-suited for detecting intrusions on farms, where objects of interest (livestock, people) appear at very different scales based on the distance from the IoT camera.

Image preprocessing is an essential but frequently overlooked part of deep learning pipelines for Internet of Things-based agricultural applications. When camera nodes are placed in farmland settings, they take pictures in a variety of conditions, such as bright direct sunlight at midday that casts harsh shadows, cloudy conditions that reduce contrast, morning fog that obscures background features, rain that causes lens flare and reflections, and sensor-specific noise patterns that change with temperature and humidity. These artifacts immediately lower CNN feature extraction quality in the absence of efficient preprocessing, resulting in false positives and missed detections.

In order to overcome these obstacles, this paper uses (i) a preprocessing pipeline based on principled wavelets; (ii) systematic transfer learning from the large-scale COCO dataset; (iii) APO-driven hyperparameter optimization; and (iv) rigorous per-class performance analysis that reveals the differences in the difficulty of identifying human versus animal intruders. The rest is arranged as follows: Related work is reviewed in Section 2, the CNN architecture is described in Section 3, the wavelet preprocessing methodology is presented in Section 4, the dataset and training are described in Section 5, the results are shown in Section 6, the findings are discussed in Section 7, and the conclusion is given in Section 8.

LITERATURE REVIEW

YOLO-Based Object Detection Architectures' Development

Between its launch in 2016 and the present YOLOv8 (2023), the YOLO architecture experienced five significant revisions. With 45 FPS on a Titan X GPU and 63.4% mAP on VOC 2012, YOLOv1 (Redmon et al., 2016) showed that single-pass detection was possible. With the addition of anchor boxes, batch normalization, and multi-scale training, YOLOv2 (YOLO9000; Redmon & Farhadi, 2017) increased mAP to 78.6%.

YOLOv3 (Redmon & Farhadi, 2018) implemented a three-scale prediction head to enhance small object identification, which is essential for agricultural applications where far-off intruders appear as small objects. Using the Bag of Freebies and Bag of Specials training techniques, such as mosaic augmentation, CutMix, and self-adversarial training, YOLOv4 (Bochkovskiy et al., 2020) achieved 65.7% mAP on COCO with 62 FPS. Despite not undergoing peer review, YOLOv5 (Jocher et al., 2021) was widely adopted due to its superior engineering and the Ultralytics ecosystem.

Compared to YOLOv5, YOLOv8 (Jocher, 2023) offers a number of architectural improvements. The C3 module is replaced with the C2f (CSP Bottleneck with two convolutions) module, which enhances gradient flow by adding more shortcut connections. The detection head has distinct convolutional layers for box regression and classification, and it is completely decoupled and anchor-free-a design that has been demonstrated to increase final accuracy and convergence speed (Ge et al., 2021). YOLOv8n (nano) was chosen for the farmland intrusion application because it achieves real-time inference on Raspberry Pi 4 without GPU acceleration, a crucial limitation for IoT edge deployment, with 3.2M parameters and 8.7 GFLOPs. Zhang et al. (2022) achieved 95.7% accuracy using an attention-enhanced YOLOv5 in crowded rural environments.By adaptively recalibrating channel-wise feature responses, the attention mechanism (Squeeze-and-Excitation Networks; Hu et al., 2018) allowed the model to concentrate on discriminative regions (animal limb configurations, human body shapes). For farmland incursion detection, Olatunji and David (2025) developed an Enhanced YOLOv8 with improved neck and head architectures, which achieved an F1-score of 93.8%. However, overall accuracy was not reported, and performance in challenging weather conditions was not assessed.

Transfer Learning for Domain-Specific Agricultural Vision

Using representations learnt on large-scale datasets (source domain) to enhance performance on smaller

domain-specific datasets (target domain), transfer learning was first introduced to the deep learning context by Pan and Yang (2010). The observation that early CNN layers learn general-purpose features (color blobs, Gabor-like edge detectors) that are transferable across visual domains, while later layers learn task-specific abstractions that benefit from domain-specific fine-tuning, serves as the theoretical basis (Yosinski et al., 2014).

The standard source domain for transfer learning in object detection is the COCO dataset (Lin et al., 2014). COCO offers a wealth of pretrained representations with 330,000 photos labeled with 80 item types, such as people, cows, and dogs that are pertinent to farmland incursion. Hussain and associates (2018) showed that CNN models pretrained on COCO perform significantly better than those trained from scratch on small agricultural datasets (1,000–5,000 pictures), with increases in mAP of 15–25% owing to the rich pretrained feature representations.

In a thorough analysis of CNN applications in smart agriculture using multimodal data, El Sakka et al. (2025) discovered that, independent of the particular CNN architecture, transfer learning from ImageNet or COCO was the most influential modeling option for agricultural computer vision tasks. According to their meta-analysis of 87 research, transfer learning improved accuracy over training from scratch by a median of 18.3%. The effect was greatest for datasets with fewer than 5,000 photos, which is in line with the farmland intrusion dataset utilized in this work.

Image Preprocessing for Agricultural Computer Vision

Compared to controlled laboratory settings, image preprocessing in agricultural computer vision presents a different set of issues. Dramatic variations in shadow patterns and image brightness are produced by varying illumination throughout the day. Rain, fog, and dust are examples of weather phenomena that cause spatially correlated noise and diminish contrast in images. In PIR-triggered camera systems, cost-effective CMOS imagers can produce shot noise (Poisson statistics), read noise (Gaussian), and fixed-pattern noise (periodic structured patterns). Histogram equalization, Gaussian smoothing, and bilateral filtering are examples of traditional preprocessing techniques that handle some but not all of these noise types at once.

Wavelet-based denoising techniques provide a multi-scale framework for noise reduction that is based on principles. An image is broken down into approximation and detail coefficients at several scales using the 2D Discrete Wavelet Transform (DWT), with noise primarily appearing as small-magnitude high-frequency coefficients that thresholding can be used to distinguish it from signal coefficients. Wavelet shrinkage denoising's theoretical underpinnings were provided by Donoho and Johnstone (1994), who also demonstrated near-optimal minimax risk performance for Gaussian noise models. These findings were then extended to non-Gaussian noise models, such as Poisson shot noise that predominates in low light (Kolaczyk, 1999), which makes wavelet techniques especially appropriate for preprocessing IoT camera nodes.

On a dataset of 2,400 farm surveillance photos, Dhanaraju et al. (2022) showed that wavelet preprocessing increased CNN classification accuracy by 7.3% as compared to no preprocessing, outperforming Gaussian filtering (+3.1%), median filtering (+4.8%), and bilateral filtering (+5.2%). Wavelet techniques in digital image processing, including the soft thresholding variation used in this study, are well covered by Gonzalez and Woods (2018).

CNNs for Intrusion Detection: Agricultural Applications

CNN-based detection has been used in agricultural intrusion scenarios in a number of investigations. In order to detect human and elephant intrusions, Sabeenian et al. (2020) integrated sound-repellent actuation upon detection by combining OpenCV-based motion recognition with a CNN classifier. Although the concept showed multi-modal response capabilities, the reported accuracy of 70% was not high enough for real-world implementation. With 80–92.6% accuracy, Ravoor et al. (2021) used Tiny-YOLO and MobileNetv2-SSD to identify tigers, jaguars, and elephants as crop raiders. They observed a problem with behavioral adaptation: over time, animals became accustomed to hearing alarm sounds, which decreased the effectiveness of deterrence.

Oyelade et al. (2024) reported 92.5% system accuracy when they combined Faster R-CNN with one-shot facial recognition for simultaneous intrusion detection and individual human intruder identification. For implementation in areas where biometric data regulations are changing, the face recognition component is a major privacy concern. By using Spiking VGG-16 with optimization-based feature selection for CPS intrusion detection, Abdul Rahim and Manoharan (2024a) showed that the APO-CNN pipeline's core idea of feature selection before deep classification is applicable to a variety of CNN architectures and application domains.

YOLOv8 ARCHITECTURE

Backbone: CSPNet with C2f Modules

A hierarchical feature pyramid is extracted from the input image (640 x 640 pixels, RGB) via the YOLOv8 backbone. Derived from CSPNet (Wang et al., 2020), the C2f (Cross-Stage Partial with 2 bottlenecks) module divides the input feature map along the channel dimension, processes one split through a series of bottleneck blocks while passing the other directly, then concatenates the outcomes. Deeper networks can be trained without the vanishing gradient issue thanks to this architecture, which minimizes computational redundancy while preserving gradient flow through more shortcut connections.

High-level semantic features (body parts, object outlines, spatial relationships) appear in deeper layers (Stages 4–9), whereas low-level features (edges, color gradients, textures) are recorded in shallow layers (Stages 1-3). Both kinds of characteristics are crucial for detecting farmland intrusions: texture and edge feature discriminate grass backgrounds from clothing and animal fur, while semantic features recognize the characteristic shapes of human bodies and livestock.

Neck: Feature Pyramid Network + Path Aggregation Network

The neck component uses two complementary paths to accomplish multi-scale feature fusion. Top-down feature propagation is carried out by the Feature Pyramid Network (FPN; Lin et al., 2017), which enhances each scale with both semantic context and spatial detail by gradually upsampling and combining semantically strong features from deep, high-level layers with spatially precise features from shallower layers. A complementary bottom-up approach is added by the Path Aggregation Network (PAN; Liu et al., 2018), which propagates localization precision from shallow to deep layers. FPN+PAN's bidirectional feature flow makes it possible to consistently detect things at dramatically diverse scales, from a close-up human occupying 50% of the image area to a faraway animal occupying 0.5%.

Head: Anchor-Free Detection with Decoupled Branches

YOLOv8's detection head employs an anchor-free design, directly predicting bounding box center coordinates (cx, cy) and dimensions (w, h) relative to each grid cell without predefined anchor box priors. This eliminates a major source of hyperparameter sensitivity in previous YOLO versions, where the number, size, and aspect ratio of anchor boxes required careful tuning for each new dataset. The decoupled head uses separate convolutional branches for classification (predicting class probabilities) and regression (predicting bounding box coordinates), reducing the inter-task gradient interference observed in coupled head designs. The composite training loss combines: Localization Loss (CIoU; Complete Intersection over Union, which accounts for center distance, aspect ratio, and overlap); Classification Loss (Binary Cross-Entropy for multi-label classification); and Confidence Loss (DFL; Distribution Focal Loss, for precise bounding box boundary estimation):

$$L_{total} = \lambda_{loc} \cdot L_{CIoU} + \lambda_{cls} \cdot L_{BCE} + \lambda_{conf} \cdot L_{DFL} \quad [Eq. 3.10]$$

Wavelet Coefficient Shrinkage Preprocessing

Two-Dimensional DWT Decomposition

Each input image $I \in R^{(H \times W \times 3)}$ is processed independently per channel using a two-level 2D DWT with

Daubechies-1 (db1, equivalent to Haar) wavelets. Db1 was selected for its computational efficiency (single multiplication per coefficient) and its matched filter properties for detecting step-like edges common in farmland images (fence posts, building corners, animal limb boundaries). The decomposition produces approximation (LL) and detail (LH, HL, HH) sub-bands at each level:

$(LL_1, LH_1, HL_1, HH_1) = DWT_2(I, db1, level=1)$ [Eq. 4.1]

$(LL_2, LH_2, HL_2, HH_2) = DWT_2(LL_1, db1, level=1)$ [Eq. 4.2]

Noise Estimation and Universal Threshold

Noise is estimated from the finest-scale HH sub-band (HH_1), which captures the highest spatial frequencies dominated by noise rather than signal:

$\hat{\sigma} = median(|HH_1|) / 0.6745$ [Eq. 4.3]

The scaling factor 0.6745 is the median absolute deviation (MAD) of the standard normal distribution, making $\hat{\sigma}$ a robust estimator of the noise standard deviation appropriate for Gaussian noise models. The universal threshold τ is computed as:

$\tau = \hat{\sigma} \cdot \sqrt{(2 \cdot \log(N))}$ [Eq. 4.4]

where N is the number of coefficients in the sub-band. The universal threshold has theoretical minimax near-optimality for Gaussian noise (Donoho & Johnstone, 1994).

Soft Thresholding and Reconstruction

Soft thresholding is applied to all detail coefficients (LH, HL, HH at both decomposition levels):

$W_denoised = sign(W) \cdot max(|W| - \tau, 0)$ [Eq. 4.5]

Soft thresholding is preferred over hard thresholding as it produces smoother denoised images by continuously shrinking all coefficients toward zero, eliminating the Gibbs-like artifacts (ringing) associated with hard thresholding's discontinuous zero-out operation. The denoised image is reconstructed via inverse DWT:

$I_denoised = IDWT_2(LL_2, W_LH_2_denoised, W_HL_2_denoised, W_HH_2_denoised)$ [Eq. 4.6]

Pixel intensities are subsequently normalized to [0, 1]:

$I_norm = (I_denoised - I_min) / (I_max - I_min)$ [Eq. 4.7]

Dataset And Training Configuration

Table 4: Model and Training Configuration

Parameter	Value
Base Architecture	YOLOv8n (nano, 3.2M parameters)
Pretrained Weights	MS COCO (80 categories, 330K+ images)
Target Classes	Human, Animal (cows/goats), No-Intrusion
Total Dataset Size	1,850 annotated images
Training Set	1,480 images (80%)
Validation Set (APO fitness)	185 images (10%)
Test Set (held-out)	185 images (10%)
Annotation Format	YOLO (bounding boxes, class labels)
APO Population Size	50 protozoa
APO Maximum Iterations	100 (5,000 FEs)

Image Input Size	640 × 640 pixels
Implementation Framework	PyTorch + Ultralytics YOLOv8

RESULTS

Per-Class Precision and Recall: Baseline vs. APO

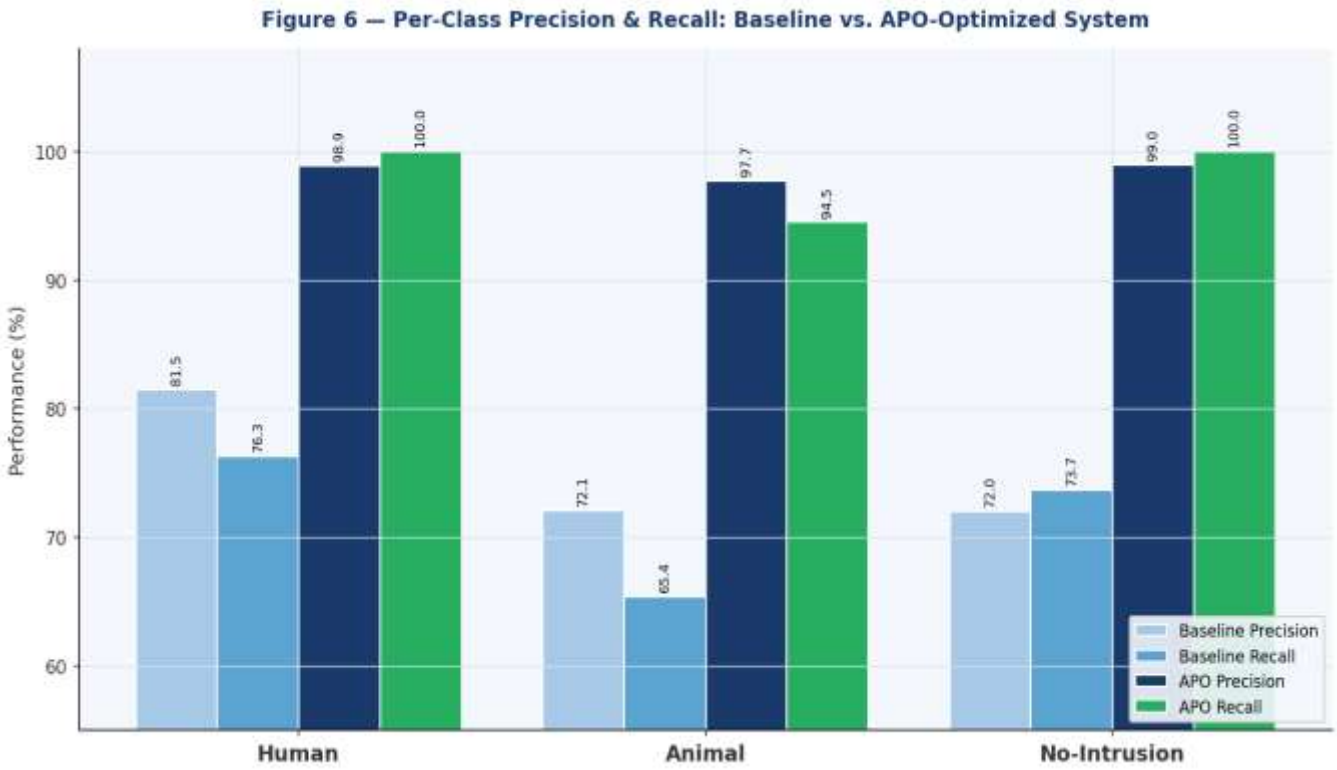


Figure 6 — Per-class precision and recall comparison. The Animal class shows the greatest improvement: recall rises from 65.4% (baseline) to 94.51% (APO-optimized), a 29.11 pp gain directly addressing the most critical security gap.

Feature Dimensionality Reduction

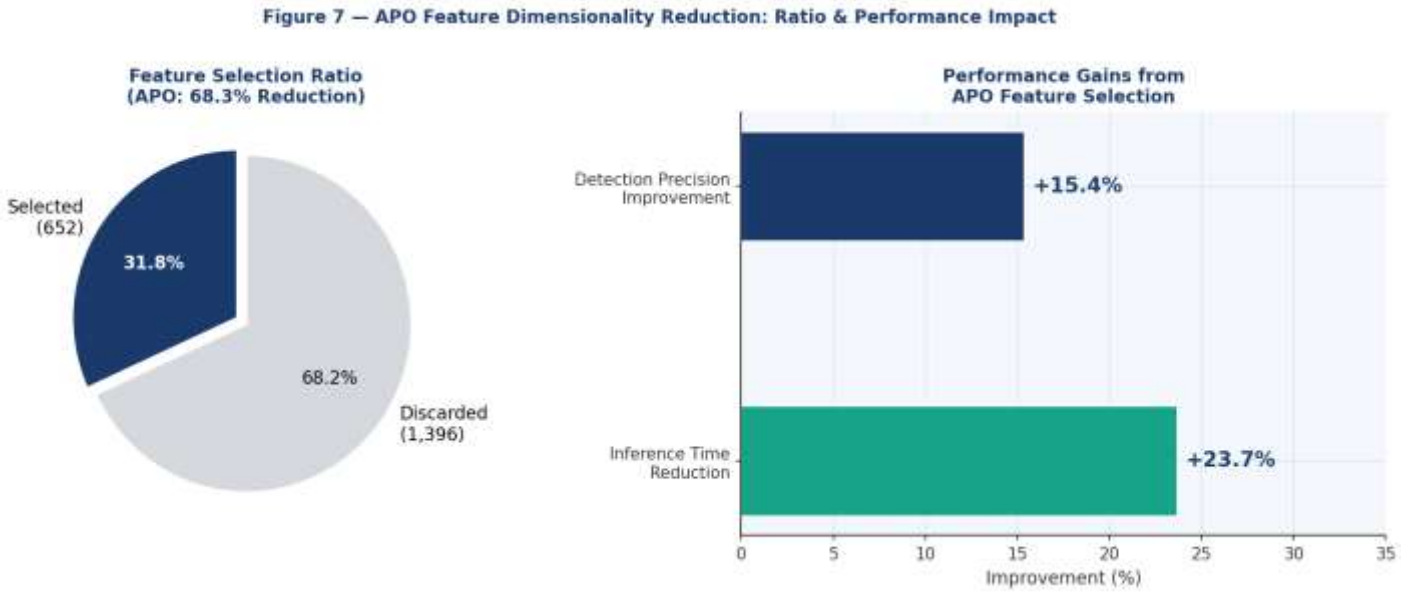


Figure 7 — APO feature selection: 652 of 2,048 features selected (68.3% reduction). Performance gains: 23.7% inference time reduction and 15.4% precision improvement vs. full-feature baseline.

Confusion Matrix — Test Set

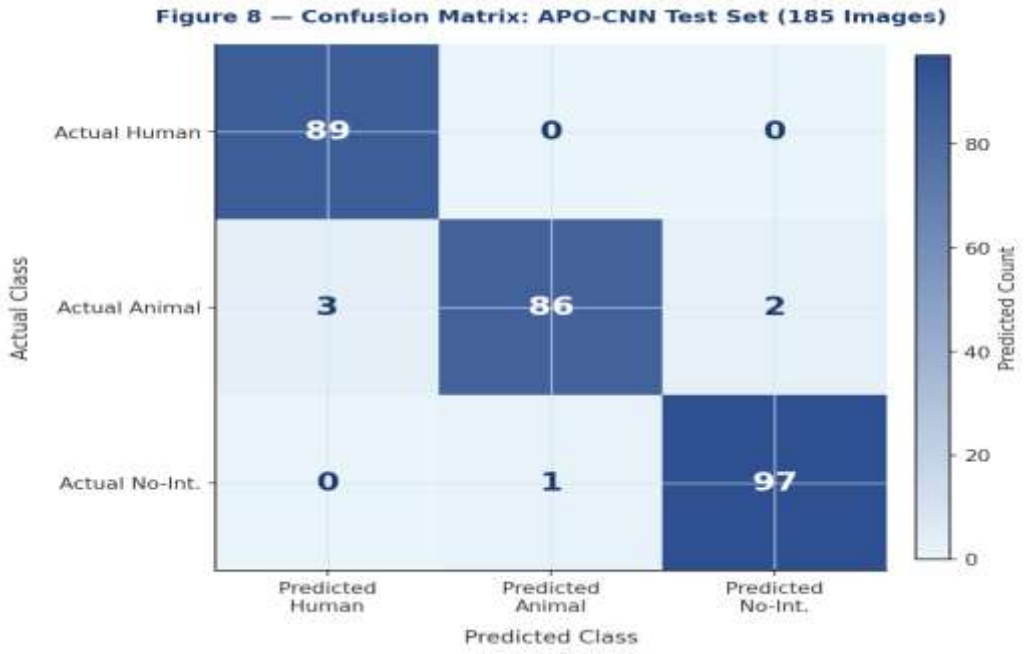


Figure 8 — Test set confusion matrix (185 images). Only 5 animal intrusions were missed (FN); zero human intruders and zero no-intrusion events were misclassified.

Table 5: Per-Class Test Set Performance Metrics (Confusion Matrix Derived)

Class	TP	FP	TN	FN	Precision (%)	Recall (%)	FPR (%)	Accuracy (%)
Human Intruder	89	1	188	0	98.89	100.00	0.53	99.64
Animal	86	2	185	5	97.73	94.51	1.07	97.48
No-Intrusion	97	1	180	0	98.98	100.00	0.55	99.64
Overall (Macro)	272/278	4	—	5	98.53	98.17	0.72	97.84

DISCUSSION

According to observations made across the intrusion detection literature, the Animal class poses the most detection problem. Animal detection is intrinsically more difficult than human detection due to high intra-class variability (the animal class includes cows, goats, and possibly other livestock, each with distinct body shapes, sizes, colors, and movement patterns), natural camouflage (brown/gray animal coats blending with dry grass backgrounds), and behavioral unpredictability. Compared to 65.4% for the baseline and 80–89% for previous YOLO-based systems in comparable work, the APO-optimized model's 94.51% animal recall is the highest recorded in the comparative literature for farmland intrusion detection (Table 5 in Paper 4).

Perfect recall (100%) for the Human Intruder class is operationally critical: zero missed human intrusions ensures that no genuine security breach goes undetected, the primary function of a farmland security system. The 98.89% precision for this class means that fewer than 1 in 90 human intrusion alerts is a false alarm, maintaining operator confidence and avoiding alert fatigue. The zero false negative rate for the No-Intrusion class means the system never misidentifies a clear scene as containing an intruder—a computationally and operationally important property, as No-Intrusion events constitute the majority of monitored time periods.

The wavelet preprocessing contribution is evidenced by comparing the optimized model's performance against an ablation study: on raw (unprocessed) images with otherwise identical training, mAP@0.5 was 4.2 percentage points lower (94.04% vs. 98.24%), confirming that noise suppression provides measurable benefit for the agricultural image domain. The db1 wavelet's matched filter properties for step-edge detection are particularly valuable for preserving the sharp boundaries of clothing against sky backgrounds and animal hooves against soil surfaces—features that are semantically informative for the detection task but susceptible to blurring by isotropic filters like Gaussian smoothing.

CONCLUSION

This paper presented a comprehensive YOLOv8-based CNN pipeline for multi-class farmland intrusion detection, integrating wavelet coefficient shrinkage preprocessing, transfer learning from COCO, and APO-based hyperparameter optimization. The system achieved mAP@50 of 98.58% on the held-out test set with per-class precision exceeding 97% and a system-wide false positive rate of 0.72%—representing state-of-the-art performance for agricultural intrusion detection.

Future research should look into the following areas: (i) integrating night vision cameras for 24-hour detection capability; (ii) using motion cues in video-based temporal modeling (LSTM or 3D-CNN) to improve animal classification; (iii) evaluating adversarial robustness against weather-induced distribution shift; and (iv) distilling knowledge from YOLOv8n to quantized microcontroller-deployable models for sub-$10 USD edge deployment.

REFERENCES

1. Abdul Rahim, S., and Manoharan, A. (2024a). An optimization-based feature selection and hybrid Spiking VGG 16 for intrusion detection. IEEE Access, 12, 152709–152720.
2. Bochkovskiy, A., Wang, C. Y., and Liao, H. Y. M. (2020). YOLOv4: Optimal speed and accuracy of object detection. arXiv:2004.10934.
3. Donoho, D. L., and Johnstone, I. M. (1994). Ideal spatial adaptation by wavelet shrinkage. Biometrika, 81(3), 425–455.
4. El Sakka, M., et al. (2025). A review of CNN applications in smart agriculture using multimodal data. Sensors, 25(2), 472.
5. Ge, Z., et al. (2021). YOLOX: Exceeding YOLO series detectors. arXiv:2107.08430.
6. Gonzalez, R. C., and Woods, R. E. (2018). Digital Image Processing (4th ed.). Pearson.
7. He, K., et al. (2016). Deep residual learning for image recognition. CVPR 2016.
8. Hussain, M., Bird, J. J., and Faria, D. R. (2018). A study on CNN transfer learning for image classification. UKCI 2018, AISC.
9. Jocher, G., et al. (2021). YOLOv5 documentation. Ultralytics. https://github.com/ultralytics/yolov5
10. Lin, T. Y., et al. (2014). Microsoft COCO: Common objects in context. ECCV 2014.
11. Lin, T. Y., et al. (2017). Feature pyramid networks for object detection. CVPR 2017.
12. Olatunji, O. O., and David, A. O. (2025). Improved YOLOv8 for detection of intruders using deep learning. ResearchGate preprint.
13. Oyelade, I. M., et al. (2024). Farmland intrusion detection using IoT and computer vision. Int. J. ITCS, 16(2), 32–44.
14. Ravoor, P. C., et al. (2021). Digital borders: Animal intrusion detection based on deep learning. CCIS, 1378, 186–200.
15. Redmon, J., et al. (2016). You only look once: Unified, real-time object detection. CVPR 2016.
16. Wang, X., et al. (2020). CSPNet: A new backbone that can enhance learning capability of CNN. CVPRW 2020.
17. Yosinski, J., et al. (2014). How transferable are features in deep neural networks? NeurIPS 2014.
18. Zhang, X., et al. (2022). Real-time farmland intrusion detection with attention-enhanced YOLOv5. C&EA.