

A Hybrid Robust Learning Framework for Multi-Scale Chicken Detection in Extreme Battery Cage Environments Using Augmented Transfer Learning

Abdul Malik I. Buna

Department of Electrical Engineering, Faculty of Engineering, Universitas Hasanuddin, Gowa, Indonesia
| Department of Informatics, Faculty of Computer Science, Ichsan Sidenreng Rappang University,
Sidenreng Rappang, Indonesia
bunaami22d@student.unhas.ac.id

Zahir Zainuddin

Department of Informatics, Faculty of Engineering, Universitas Hasanuddin, Gowa, Indonesia
zahir@unhas.ac.id (corresponding author)

Syafuruddin Syarif

Department of Electrical Engineering, Faculty of Engineering, Universitas Hasanuddin, Gowa, Indonesia
syafuruddin.s@eng.unhas.ac.id

Received: 15 January 2026 | Revised: 4 April 2026 | Accepted: 17 April 2026

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.17563>

ABSTRACT

Multi-scale chicken detection in battery-cage farms is a challenging computer vision problem due to the severe occlusion from cage bars, low and inconsistent illumination, and high object density. This study proposes a hybrid robust learning framework based on augmented transfer learning to improve multi-scale chicken detection. The detector is built on Faster Region-based Convolutional Neural Networks (R-CNN) with a ResNet-50-FPN backbone. It is enhanced by an early-dilated convolutional block and a regularized custom head incorporating dropout and batch normalization. Transfer learning is applied through selective fine-tuning, in which only backbone stages 3–4 are retrained while earlier layers are frozen to preserve general representations. Four training schemes were evaluated: (a) baseline without augmentation, (b) online dynamic augmentation, (c) offline augmentation using Albumentations, and (d) hybrid offline+online augmentation. The augmentation pipeline includes rotation, horizontal flip, brightness/contrast adjustment, Gaussian blur, gamma correction, and resizing to 512×512, with COCO-format bounding boxes synchronized during transformation. Experiments using COCO evaluation show that the hybrid scheme achieves the best overall in-domain performance, reaching $mAP@0.5 = 0.880$, $mAP@[0.50:0.95] = 0.414$, and $AR@100 = 0.483$, corresponding to an absolute gain of 0.098 in $mAP@0.5$ over the offline-only scheme. Despite these gains, small-object detection remains unresolved, indicating the need for higher-resolution feature representations and anchor/grid redesign for micro-scale and heavily occluded instances. Overall, the findings indicate improved detection performance and in-domain generalization under naturally challenging battery-cage conditions. However, no dedicated corruption-based robustness benchmark or stress test was conducted; therefore, the results should not be interpreted as formal evidence of robustness to explicit illumination, blur, or noise perturbations.

Keywords-object detection; faster R-CNN; hybrid augmentation; dilated convolution; selective fine-tuning; battery cage chicken; multi-scale; computer vision

I. INTRODUCTION

Improving animal welfare while minimizing losses in poultry breeding systems depends heavily on the ability to detect and provide early warnings of infectious diseases. Conventional

disease-control approaches often fall short, resulting in high mortality and reduced productivity [1]. Authors in [1] proposed a sound-based health-monitoring system using Mel-Frequency Cepstral Coefficients (MFCC) features and a Convolutional Neural Network (CNN) that can distinguish healthy from sick

chickens with 94.59% accuracy, 96% precision and recall, and 96% F1-score [1]. Building on these findings, the current study extends automation to the domain of computer vision to detect chickens in battery cages, thereby enabling the integration of visual and health monitoring.

The adoption of automation technologies is pivotal for enhancing productivity while simultaneously maintaining and improving quality [2, 3]. Modern poultry farming, especially battery-cage systems for laying hens, increasingly requires reliable automated monitoring to support animal welfare and enable early risk detection at a commercial scale [4]. In practice, routine tasks such as counting hens remain labor-intensive, time-consuming, and error-prone, motivating the development of computer vision-based monitoring systems that operate reliably in real farm environments [2].

Battery cage imagery poses challenges that differ substantially from those in conventional object detection benchmarks. These include severe occlusion from metal cage bars, inconsistent illumination (dim to harsh artificial lighting), high object density with frequent overlap, and pronounced intra-image scale variation as birds move within confined spaces [5]. Therefore, it is necessary to increase detection performance under extreme conditions.

Research on poultry management and disease detection has been widely carried out [6-8]. Transfer learning mitigates annotation scarcity by adapting pretrained models to agricultural applications [9, 10], yet conventional fine-tuning alone is frequently insufficient under extreme battery cage conditions where occlusion, clutter, and multi-scale variability dominate failure modes [5]. Data augmentation is critical for improving generalization in computer vision [11] and agricultural computer vision [12], and augmentation-enhanced transfer learning has shown benefits in related tasks [13]. However, prior studies often emphasize either offline dataset expansion or online on-the-fly augmentation; a comparative evaluation of both within a single controlled protocol is less frequently reported for extreme battery-cage poultry detection [5, 12, 14].

Recent surveys confirm that deep-learning detectors (e.g., Faster Region-based Convolutional Neural Networks (R-CNN), You Only Look Once (YOLO), Single Shot Detector (SSD)) remain strong baselines across diverse applications and benchmarks [15]. Their accuracy often degrades in poultry monitoring due to domain shift relative to general datasets such as Common Objects in Context (COCO) and Pascal Visual Object Classes (VOC) [4]. Moreover, architectural adaptations (e.g., receptive-field enhancements and customized heads) are often evaluated as single-aspect improvements rather than being integrated with augmentation design and selective fine-tuning to address domain-specific multi-scale challenges [13, 16, 17].

To bridge this gap, we propose a Hybrid Robust Learning framework for multi-scale chicken detection in challenging battery-cage imagery. The proposed framework integrates hybrid augmentation, architectural adaptation, and selective fine-tuning to improve detection performance under naturally difficult visual conditions. First, it employs a hybrid augmentation pipeline that combines offline pre-generated samples with online, on-the-fly transformations during training,

thereby enriching data diversity while preserving dynamic visual perturbations. Second, it adopts a modified Faster R-CNN detector built upon a ResNet-50 backbone with a Feature Pyramid Network (FPN), further enhanced by an early dilated convolution block and a regularized detection head to improve generalization. Rather than claiming formal robustness to explicit perturbations, the proposed approach is evaluated under a controlled experimental protocol that uses COCO-style metrics and object-scale analysis. Performance is explicitly reported within battery-cage environments, allowing the scope of applicability and domain-specific limitations to be clearly articulated.

Each considered training scheme was evaluated using the same COCO-based protocol (mAP, AR) and a scale-wise analysis (small/medium/large), enabling a controlled assessment of the effects of augmentation strategies and their interactions with architectural adaptation and selective fine-tuning. This study addresses three research questions: (1) to what extent hybrid augmentation improves multi-scale chicken detection accuracy, (2) whether architectural adaptation and selective fine-tuning improve in-domain generalization under challenging battery-cage conditions, and (3) which strategy provides the most practical performance-generalization trade-off for deployment in vision-based livestock monitoring systems. The contributions include: a controlled comparison of offline/online/hybrid augmentation for battery-cage chicken detection; a modified Faster R-CNN design incorporating dilated convolution, a regularized detection head, and selective fine-tuning optimized for poultry monitoring; and a comprehensive COCO-based multi-scale evaluation that yields deployment-oriented insights for precision livestock farming [18].

II. METHODOLOGY

A. Model Architecture

The proposed detector is based on Faster R-CNN with a ResNet-50 backbone and an FPN for multi-scale representation. To better handle dense packing, partial occlusion, and large-scale variation in battery-cage imagery, three targeted modifications are introduced: (i) an early dilated convolution block, (ii) a regularized detection head with dropout and batch normalization, and (iii) selective fine-tuning of the backbone.

Input RGB images of $512 \times 512 \times 3$ size were first processed by a 3×3 dilated convolution layer (dilation rate = 2, padding = 2), followed by Batch Normalization and ReLU activation. This early dilated block is intended to enlarge the receptive field without reducing spatial resolution, enabling the network to capture broader contextual cues useful in crowded and partially occluded scenes. The resulting features were then forwarded to the ResNet-50-FPN backbone, where pyramid levels P2-P6 support detection across different object scales.

To adapt the pretrained detector to the target domain while preserving transferable low-level visual patterns, stages 1-2 of the backbone were frozen, whereas stages 3-4 and the FPN were fine-tuned. This selective fine-tuning strategy was motivated by the fact that lower layers generally encode generic features such as edges and textures. In contrast, deeper layers are more task-

specific and therefore more relevant to adapting to battery-cage chicken imagery.

To reduce overfitting under limited data and strong appearance variability, the detection head incorporated Dropout ($p = 0.3$) followed by BatchNorm2d. Dropout was used to limit feature co-adaptation, while batch normalization stabilizes optimization during training. Region proposals were generated

by a Region Proposal Network (RPN) using anchor sizes $\{32, 64, 128, 256, 512\}$ and aspect ratios $\{0.5, 1.0, 2.0\}$, followed by ROI Align and ROI Head prediction for classification and bounding-box regression. These architectural choices should be interpreted as theory-guided design decisions for the current study rather than individually ablation-optimized settings. Figure 1 summarizes the overall architecture.

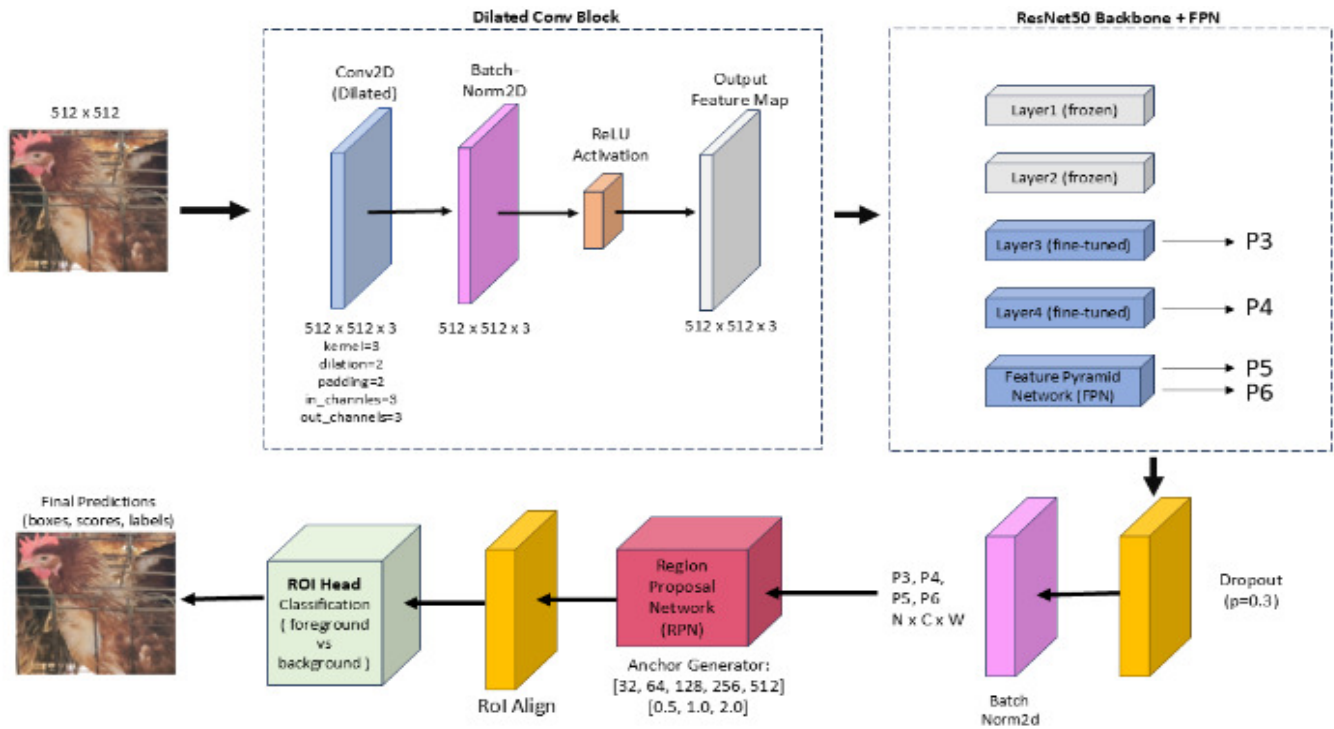


Fig. 1. Proposed architecture.

B. Augmentation Strategy and Experimental Schemes

To improve detection performance under complex visual conditions, this study systematically evaluates the effect of offline, online, and hybrid augmentation strategies on the proposed modified Faster R-CNN. Four training schemes were defined (Table I) to isolate the contribution of each augmentation modality and quantify its impact on accuracy and generalization within the battery-cage domain. In this article, the term Hybrid Robust Learning refers to the overall framework design rather than to a formally benchmarked robustness setting.

TABLE I. PROPOSED TRAINING SCHEMES

Scheme	Label	Aug. Mode	When applied	Description
A	Baseline	None	-	Baseline without augmentation
B	Online augmentation	Online	During training	Online dynamic augmentation using augmentations during training
C	Offline augmentation	Offline	Before training	Offline augmentation with a pre-generated augmented training dataset.
D	Hybrid augmentation	Offline + Online	Before + during training	A hybrid combination of offline and online augmentation.

- Scheme A (Baseline) serves as the control condition, in which the model is trained on the original dataset without augmentation, thereby establishing the reference performance under limited data diversity.
- Scheme B (Online Augmentation) applies augmentation on the fly during training, introducing stochastic transformations per batch, including horizontal flip, brightness/contrast adjustment, Gaussian blur, rotation ($\pm 10^\circ$), gamma correction, and resizing to 512x512. This setting is intended to mimic real-world variability and improve generalization under unpredictable distortions [19, 20].
- Scheme C (Offline Augmentation) applies the same transform family during training to produce a fixed-augmented dataset. Although offline augmentation provides less stochastic variability than online augmentation, it improves dataset diversity while shifting the augmentation cost to a preprocessing step.
- Scheme D (Hybrid Augmentation) combines offline pre-augmentation with online augmentation during training, leveraging both fixed diversity and dynamic perturbations [14]. This dual exposure is hypothesized to improve learning

stability and in-domain generalization in highly occluded and variable settings, such as poultry cages [5].

Figure 2 summarizes the experimental workflows: Scheme A trains directly on the original dataset, Scheme B augments samples dynamically during training, Scheme C trains on a pre-augmented dataset, and Scheme D integrates both offline-

augmented inputs and online augmentation within the training loop. By comparing these four schemes, this study provides controlled evidence on how augmentation modality influences detection performance and generalization in agricultural computer vision tasks, where operating conditions frequently deviate from those in controlled environments.



Fig. 2. Experimental schematic flow.

C. Dataset

The dataset consists of RGB images of laying hens captured directly in battery-cage environments at a poultry farm in Kotamobagu, North Sulawesi Province, Indonesia, in September 2024. All images were manually annotated using Label Studio and exported in COCO format. The battery-cage setting was intentionally selected because it represents a challenging detection environment characterized by high object density, large intra-image scale variation, inconsistent illumination, and severe occlusion caused by cage bars and overlapping birds.

To ensure reproducibility and prevent data leakage, the original non-augmented images were first partitioned at the image level into non-overlapping training, validation, and test

sets before any offline augmentation was applied. The original training set contains 171 images with 4,913 annotations, the validation set contains 68 images with 1,934 annotations, and the test set contains 50 images, as summarized in Table II. The validation and test sets were kept fixed across all experimental schemes to ensure a fair comparison.

Schemes A and B used the original training split without offline expansion. In contrast, Schemes C and D used an offline-augmented version of the training split, increasing the training data to 684 images with 19,601 annotations. Offline augmentation was generated by creating transformed variants of each training image using an Albumentations-based pipeline, while the corresponding bounding boxes were synchronously

updated to preserve annotation integrity. Importantly, all offline-augmented samples were restricted to the training partition. No validation or test images, nor any augmented derivatives of them, were included in model training.

TABLE II. DATASET COMPOSITION AND DATA PARTITIONS USED

Scheme	Training Images	Training Annotations	Validation Set	Test Set
A & B	171	4,913	68 images / 1,934 annotations	50 images
C & D	684	19,601	68 images / 1,934 annotations	50 images

The training set was used for model learning, while the fixed validation set was used for quantitative evaluation and comparison across schemes. All COCO-based metrics reported in Table III, along with the related quantitative analyses, were obtained from the validation set. The test set was kept fully unseen during training and validation and was used for final qualitative inspection of detection behavior on unseen images.

The offline augmentation pipeline includes horizontal flip, random brightness/contrast adjustment, rotation ($\leq 10^\circ$), shift-scale - rotate (translation and scaling), and random gamma adjustment, each applied with predefined probabilities to increase appearance diversity. Object bounding boxes were automatically transformed and updated to remain consistent with the augmented images, ensuring annotation integrity after geometric transformations. Figure 3 presents representative examples. Offline augmentation was applied before training to expand the dataset, whereas online augmentation introduced additional variability at each epoch. All augmented images and updated annotations were consolidated into a COCO-format JSON file and used directly in the training pipeline.

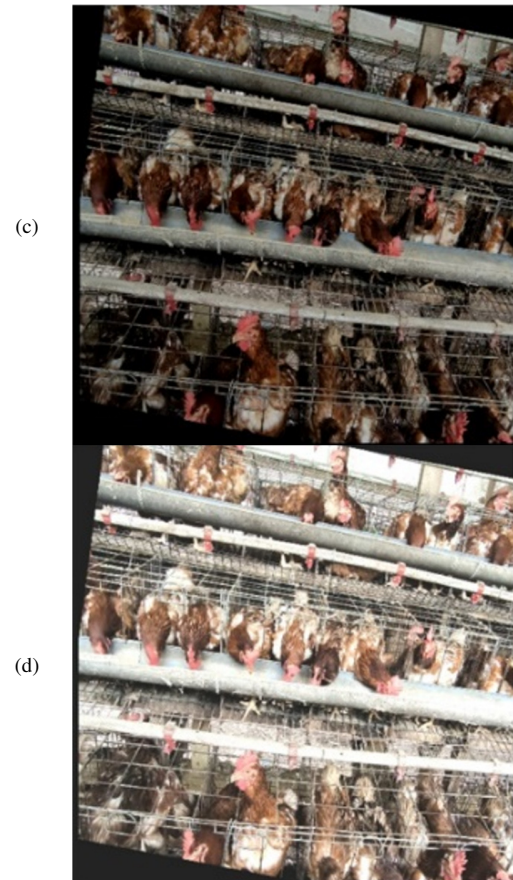
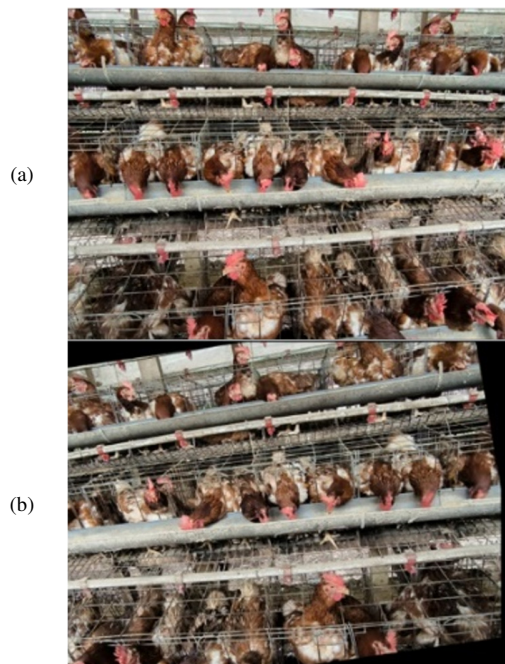


Fig. 3. Image augmentation examples. (a) Original image, (b) offline-augmented example, (c)-(d) online-augmentation examples generated dynamically during training.

By expanding training diversity through offline augmentation and generating additional variations via online augmentation, the proposed data design aims to represent better appearance shifts commonly observed in battery-cage imagery. Consequently, this dataset setup supports a controlled and reliable assessment of detection performance and multi-scale behavior across all Schemes A–D.

D. Training Procedure

Training was conducted for 50 epochs with the Stochastic Gradient Descent (SGD) optimizer (learning rate = 0.0025, momentum = 0.9, weight decay = 0.0005), with a warm-up phase applied during the first five epochs. The batch size was set to 2. All experiments were run on a Compute Unified Device Architecture (CUDA)-enabled Graphics Processing Unit (GPU), with regular checks for bounding box stability, empty-box detection, and validation of non-finite loss.

Figure 4 shows the training loss curves for all schemes throughout 50 epochs. Models C and D, which use offline and hybrid augmentations, respectively, demonstrate faster, more stable convergence than Models A and B. However, Model C exhibits a noticeably lower training loss than the other models, especially after epoch 20. This rapid decrease and stabilization in loss, while initially indicating efficient learning, may instead reflect stronger fitting to the training distribution. However, this

does not, by itself, confirm overfitting without complementary validation-loss evidence, in which the model fits the training data too closely and generalizes poorly to unseen data. In contrast, Model D achieves consistently low loss while converging more gradually, suggesting better generalization from the hybrid augmentation approach. Model B shows the slowest loss reduction, indicating that online augmentation alone is less effective in this context.

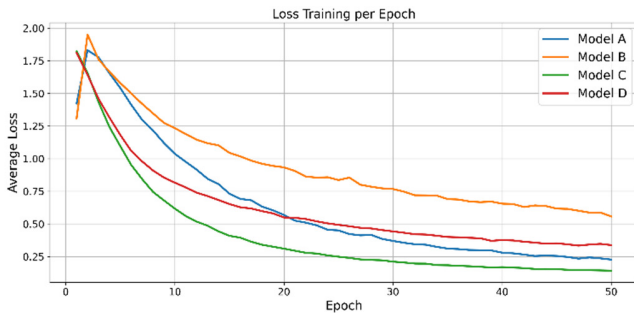


Fig. 4. Training loss per epoch.

E. Evaluation

Evaluation was conducted using COCOeval, which reports mAP@0.5, mAP@[0.5:0.95], mAP@0.75, and Average Recall (AR) at maxDets = 1, 10, and 100. Performance was also analyzed by object size (small, medium, large), average confidence score, and the number of detections exceeding confidence thresholds of 0.90 and 0.95. To ensure a consistent in-domain assessment, the same COCO-based protocol was applied across all training schemes.

As shown in Figure 5, the four schemes exhibit clear differences in mAP and AR, confirming that the augmentation strategy substantially affects detection performance in challenging battery-cage scenes. The baseline model (A) provides moderate performance, achieving mAP@0.5 = 0.727, mAP@[0.5:0.95] = 0.317, mAP@0.75 = 0.207, and AR@100 = 0.394. In contrast, the online augmentation scheme (B) underperforms the baseline across the main COCO metrics, with mAP@0.5 = 0.675, mAP@[0.5:0.95] = 0.229, mAP@0.75 = 0.066, and AR@100 = 0.318, suggesting that online augmentation alone is insufficient for this dataset.

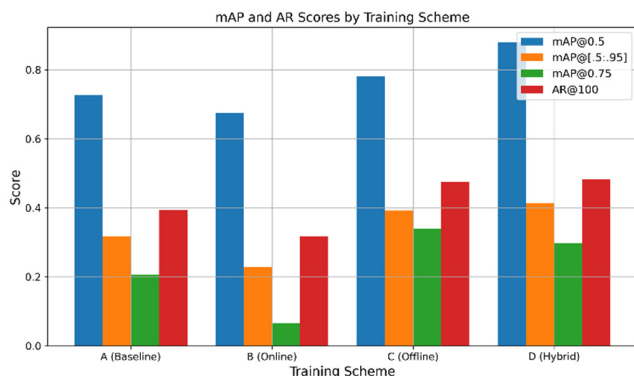


Fig. 5. Comparative results of mAP and AR scores.

The offline augmentation scheme (C) improves performance over the baseline, yielding mAP@0.5 = 0.782, mAP@[0.5:0.95] = 0.393, mAP@0.75 = 0.341, and AR@100 = 0.476. It also achieves the highest average confidence (0.8910), the highest number of high-confidence detections (Detect>0.90 = 1,531; Detect>0.95 = 1,459), and the lowest final loss (0.1418). The hybrid augmentation scheme (D) delivers the best overall in-domain performance on the main detection metrics, achieving mAP@0.5 = 0.880, mAP@[0.5:0.95] = 0.414, and AR@100 = 0.483, while also achieving the strongest performance on medium- and large-object classes. These results indicate that combining offline and online augmentation is the most effective strategy for improving overall detection performance in naturally challenging battery-cage conditions. However, the findings should not be interpreted as formal evidence of robustness to explicit corruptions.

As depicted in Figure 6, the hybrid augmentation model (scheme D) successfully detects nearly all chickens in the battery cage, achieving high confidence scores and accurate localization, even under severe occlusion and challenging backgrounds. This finding underscores the potential of multi-level augmentation strategies and transfer learning to improve object detection performance in naturally challenging real-world environments.

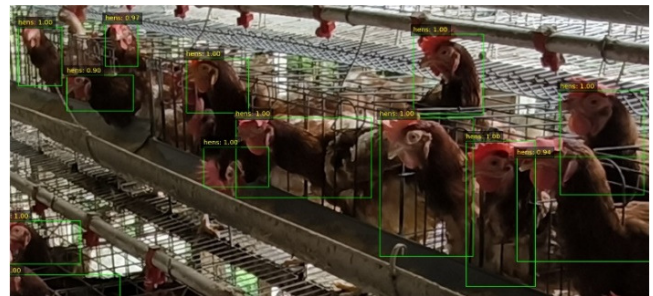


Fig. 6. Visualization of object detection results using the Hybrid Augmentation scheme on chickens in a battery cage environment. Most objects are accurately detected with high confidence, even under challenging conditions such as occlusion and dense crowds.

A key contribution of this study is the unified evaluation of augmentation strategy and detector adaptation within a single battery-cage chicken detection protocol. The results show that synchronized hybrid augmentation, combined with architectural adaptation and selective fine-tuning, improves overall in-domain detection performance under dense, occluded, and visually variable farm conditions.

III. RESULTS AND DISCUSSION

A. Quantitative Evaluation Results

Table III summarizes the COCO-based evaluation across the four training schemes and shows clear performance differences driven by augmentation modality. Overall, Scheme D (Hybrid: Offline+Online) achieves the best overall in-domain performance on the main detection metrics, indicating that combining offline and online augmentation is the most effective strategy for improving overall detection accuracy and recall in challenging battery-cage scenes.

Scheme C (Offline-only) ranks second overall and remains strong on several secondary indicators. It achieves the highest mAP@0.75, the highest average confidence score, the highest number of high-confidence detections, and the lowest final loss. These results suggest that offline augmentation yields stable and confident detections, although its overall mAP@0.5 and AR@100 remain slightly lower than those of the hybrid scheme. Scheme A (Baseline) provides moderate reference performance. In contrast, Scheme B (Online-only) underperforms the baseline across all main COCO metrics. This indicates that online augmentation alone is insufficient for this dataset and does not adequately compensate for severe occlusion, clutter, and appearance variability in battery-cage imagery.

TABLE III. EXTENDED EVALUATION RESULTS

Scheme	Augmentation	mAP @0.5	mAP @0.5:0.95	mAP @0.75	Total Detect	Conf. Avg	Detect > 0.90	Detect > 0.95	Final Loss	AP Small	AP Medium	AP Large	AR @100
A	Baseline	0.727	0.317	0.207	2,277	0.7989	1,107	990	0.2266	-1.000	0.062	0.320	0.394
B	Online	0.675	0.229	0.066	3,082	0.5951	891	661	0.5570	-1.000	0.074	0.231	0.318
C	Offline	0.782	0.393	0.341	2,052	0.8910	1,531	1,459	0.1418	0.000	0.105	0.399	0.476
D	Hybrid	0.880	0.414	0.298	2,258	0.8217	1,505	1,396	0.3376	0.000	0.141	0.417	0.483

Note: AP_small follows the COCO definition (area < 32² pixels). A value of -1.000 indicates that the metric slice was undefined in COCOeval, whereas 0.000 indicates that the slice was valid but the detector failed to achieve successful small-object matches.

B. Evaluation by Object Scale

Figure 7 reports AP by object scale (small/medium/large) across all considered training schemes. Small-object detection remains unresolved in all schemes, with AP_Small remaining at zero or negative, indicating persistent failures to detect small or heavily occluded chickens. In contrast, medium and large objects show consistent gains as augmentation effectiveness increases. The best performance is achieved by Scheme D (Hybrid) with AP_Medium = 0.141 and AP_Large = 0.417, improving over the baseline (0.062 and 0.320). Overall, hybrid augmentation substantially improves medium- and large-scale detection, but additional strategies are needed to enhance sensitivity to small objects under extreme conditions.

The persistent zero/invalid AP_small values require a more careful interpretation. In the COCO evaluation, the small-object regime corresponds to instances with a bounding-box area below 32² pixels, and a score of -1.000 should be interpreted as an undefined metric slice rather than negative performance. In this study, this behavior is best explained by the combined effect of dataset and model constraints. First, battery-cage scenes contain severe occlusion, dense object packing, and frequent truncation by cage bars, so many distant chickens appear only as tiny visible fragments. Second, all images were resized to 512×512, which further reduces the pixel support of already small instances. Third, the RPN anchor configuration starts at 32 pixels, which is suitable for general multi-scale detection but is less sensitive to micro-scale or heavily truncated targets. Under these conditions, even small localization errors can cause substantial IoU degradation in the COCO small-object slice. Therefore, the AP_small results indicate a genuine limitation of the current operating regime for microscale chickens, rather than a contradiction of the gains observed for medium- and large-scale objects. This also explains why hybrid augmentation improves overall mAP and AR, while small-object detection remains unresolved.

Scale-wise analysis further shows that the main gains are concentrated on medium- and large-sized objects. AP_Medium increases from 0.062 in Scheme A to 0.141 in Scheme D, while AP_Large improves from 0.320 to 0.417. However, small-object detection remains unresolved: AP_Small is undefined (-1.000) in Schemes A–B and 0.000 in Schemes C–D. Therefore, although the hybrid strategy provides the strongest overall quantitative performance, additional measures are still required to improve sensitivity to micro-scale and heavily occluded chickens. These findings should be interpreted as evidence of improved in-domain detection performance and generalization, rather than formal robustness to explicit corruptions.

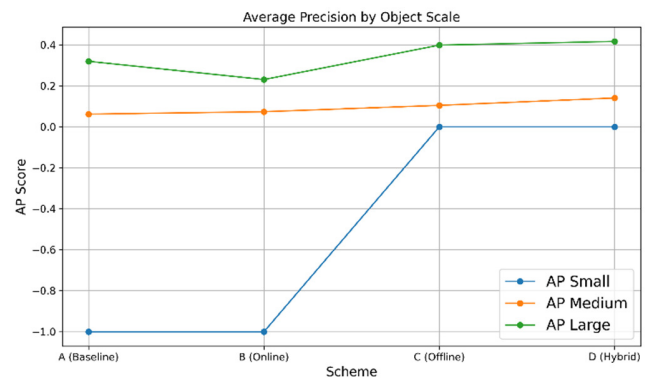


Fig. 7. AP score by object scale.

C. Qualitative Evaluation Results

Figure 8 provides qualitative comparisons of three representative battery-cage scenes. Scheme D (Hybrid) consistently yields the most complete detections, recovering more chickens in cluttered backgrounds and under partial occlusion. At the same time, Scheme C (offline augmentation) generally improves over Scheme A (Baseline) and Scheme B (online augmentation) by reducing missed detections and false positives. Nonetheless, consistent with the quantitative analysis, all schemes still miss small or heavily occluded chickens, and occasional false positives occur when cage structures or background patterns are misclassified, most frequently in the baseline and less so in the hybrid model. Overall, the visual results corroborate the COCO metrics: hybrid augmentation improves detection performance and in-domain generalization, yet small-object/occlusion failures remain a key barrier for real-world deployment.

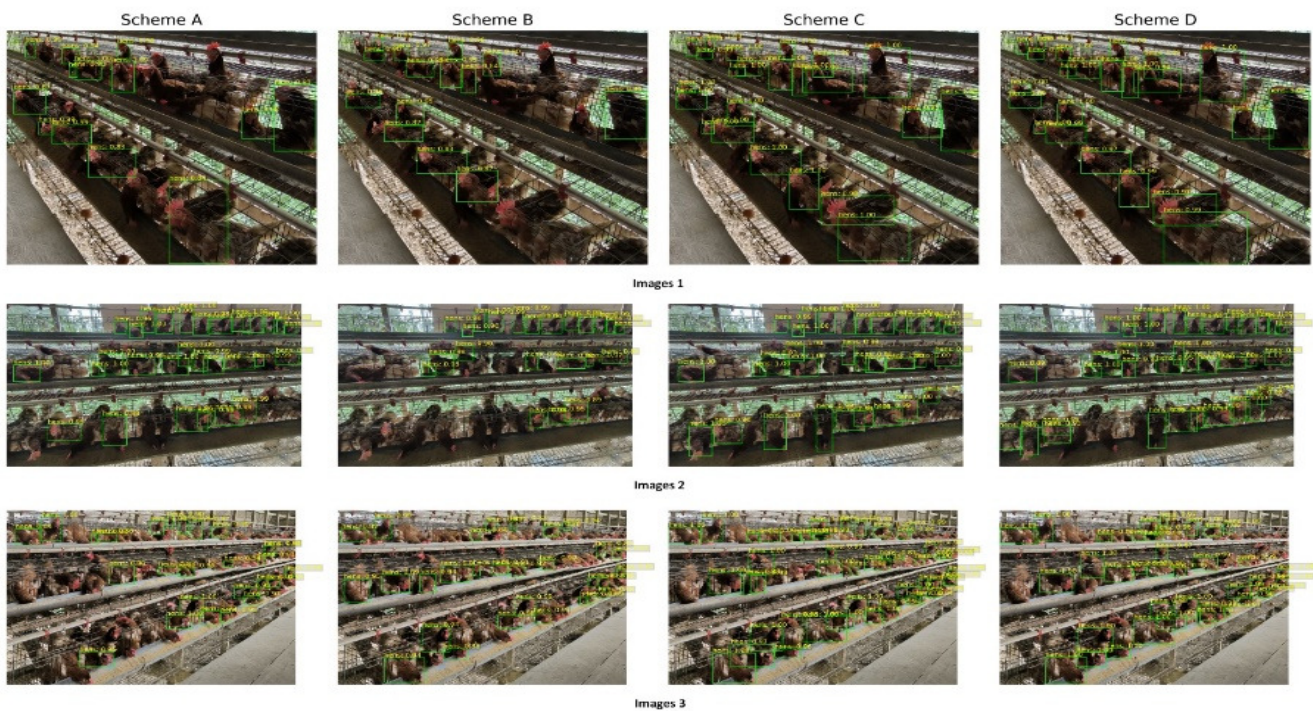


Fig. 8. Example of detection results on challenging battery cage test images using different training schemes (A: Baseline, B: Online-only, C: Offline-only, D: Hybrid). Each row corresponds to a different test image, and each column displays the results for a specific scheme. The hybrid scheme (D) detects more chickens, especially in crowded and occluded scenes, while all schemes still struggle to detect small or heavily occluded chickens.

D. Discussion

Collectively, the results show that the design of augmentations and architectural adaptations substantially influences the performance of multi-scale chicken detection in extreme battery-cage environments. Among the four evaluated schemes, the hybrid strategy (Scheme D) delivers the best overall in-domain performance, achieving the highest $mAP@0.5$, $mAP@[0.5:0.95]$, and $AR@100$. This indicates that combining offline and online augmentation is more effective than relying on either strategy alone in densely packed, occluded, and visually variable poultry scenes.

The comparative results also reveal a more nuanced pattern across schemes. Scheme C remains highly competitive and performs best on several secondary indicators, including $mAP@0.75$, average confidence, number of high-confidence detections, and final loss, suggesting that offline augmentation yields stable, confident predictions through repeated exposure to a fixed yet expanded training distribution. In contrast, Scheme B underperforms even the baseline across all principal COCO metrics, indicating that online augmentation alone is insufficient for the current dataset and training setup.

The scale-wise analysis further clarifies the strengths and limitations of the proposed framework. Performance gains are concentrated mainly on medium- and large-sized chickens, while small-object detection remains unresolved across all schemes. This limitation is consistent with the dataset and model configuration, including severe occlusion, dense object packing, truncation by cage bars, image resizing to 512×512 , and the current anchor design, all of which reduce sensitivity to micro-scale instances. Under these conditions, even minor localization

errors can substantially reduce IoU for very small objects, making AP_{small} particularly difficult to improve.

The qualitative results are consistent with the quantitative findings. Scheme D produces more complete detections in crowded, visually complex battery-cage scenes and reduces missed detections compared with the baseline and the online-only setting. However, all schemes still struggle with small or heavily occluded chickens, and occasional false positives remain visible in cage structures and background regions. These results indicate that the proposed framework improves overall in-domain detection performance, while small-object sensitivity remains a key limitation for practical deployment.

A major limitation of this study is that robustness was not evaluated using a dedicated corruption benchmark or controlled stress testing. Although the dataset naturally contains challenging visual factors such as occlusion, clutter, density, and illumination variation, the experiments are restricted to a single in-domain battery-cage setting. Therefore, the findings should be interpreted as evidence of improved detection performance and in-domain generalization under naturally challenging farm conditions, rather than as formal robustness to explicit perturbations such as synthetic blur, noise, or illumination shifts. Future work should prioritize improving sensitivity to small and heavily occluded chicken groups, as well as corruption-based evaluation and broader out-of-domain validation.

E. Summary of Key Findings

Table IV summarizes the main findings across the four training schemes. Scheme D provides the best overall in-domain performance on the principal detection metrics, whereas Scheme

C remains superior on several confidence- and localization-related indicators. Scheme B performs the worst overall and does not surpass the baseline Scheme A. Across all schemes, small-object and heavily occluded detection remains the primary limitation.

TABLE IV. SUMMARY OF KEY FINDINGS

Scheme	Findings
Scheme A	Moderate baseline performance with limited recall in challenging battery-cage scenes.
Scheme B	Underperforms the baseline across the main COCO metrics, indicating that online augmentation alone is insufficient in this setting.
Scheme C	Strong overall performance with the highest mAP@0.75, highest average confidence, highest number of high-confidence detections, and lowest final loss, but still below Scheme D on the main overall metrics.
Scheme D	Best overall in-domain performance, achieving the highest mAP@0.5, mAP@[0.5:0.95], AR@100, and the strongest medium- and large-object detection performance.

F. Practical Implications

Hybrid training (Scheme D) supports practical deployment for counting and continuous monitoring in battery-cage farms using existing CCTV/static cameras, enabling earlier detection of crowding/visibility issues and reducing manual workload. The key barrier remains handling small/heavily occluded chicken groups and future work should prioritize higher-resolution feature representations (e.g., finer pyramid levels/BiFPN), anchor/stride redesign or tiling, and small-object/occlusion-targeted augmentation. The framework is transferable to other dense-object monitoring domains (e.g., livestock behavior, warehouses, surveillance).

IV. CONCLUSION

This study introduced a Hybrid Robust Learning framework for multi-scale chicken detection in extreme battery-cage environments by combining a modified Faster R-CNN with augmented transfer learning across four training schemes. Among them, the hybrid strategy (Scheme D) achieved the best overall in-domain performance, reaching mAP@0.5 = 0.880, mAP@[0.5:0.95] = 0.414, and AR@100 = 0.483, with an absolute gain of 0.098 in mAP@0.5 over the offline-only scheme. These results indicate that combining offline and online augmentation effectively improves overall detection accuracy and recall in challenging battery-cage scenes.

However, the performance gains are concentrated mainly on medium- and large-sized objects, while small-object detection remains unresolved. AP_{small} is undefined in Schemes A-B and remains 0.000 in Schemes C-D, indicating persistent limitations for micro-scale and heavily occluded instances. Overall, the findings demonstrate improved detection performance and in-domain generalization under naturally challenging battery-cage conditions. However, because no dedicated corruption-based robustness benchmark or explicit stress test was conducted, the results should not be interpreted as formal evidence of robustness to illumination, blur, or noise perturbations. Future work should prioritize sensitivity to small objects, corruption-based evaluation, and broader validation beyond the current domain setting.

V. FUTURE WORK

This research will be continued through several strategic directions:

1. Benchmarking against YOLOv8/v11, FCOS, and SSD.
2. Improving multi-scale features via (Bi)FPN (Bidirectional Feature Pyramid Network) or ASFF (Adaptive Spatial Feature Fusion).
3. Redesigning anchors/grids for dense small objects.
4. Adopting small-object-focused augmentation (e.g., Copy-Paste, mosaic, targeted oversampling).
5. Validating performance on live video streams for real-time field deployment.

VI. CLOSING REMARKS

Overall, this study presents a practical augmentation-based training framework for chicken detection in challenging battery-cage environments. The main contribution is to show that combining hybrid augmentation, targeted architectural adaptation, and selective fine-tuning can improve overall in-domain detection performance under naturally complex farm conditions. At the same time, the findings also highlight that small-object and heavily occluded detection remains a major unresolved challenge, which should guide future model refinement and evaluation beyond the current setting.

DECLARATION OF COMPETING INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

ACKNOWLEDGMENT

This work is part of the AI-based precision livestock farming program at Universitas Hasanuddin (CBS Laboratory). The authors thank the annotation team, laboratory staff, and farming partners for the technical support and data provision. We also thank the Beasiswa Pendidikan Indonesia (BPI) for their financial assistance.

DATA AVAILABILITY STATEMENT

The dataset generated and analyzed in the current study is publicly available at [21].

AI USE AND DECLARATION OF GENERATIVE AI USE

Generative AI tools were used solely to assist with language refinement and readability. All scientific content, analyses, interpretations, and conclusions were reviewed and validated by the authors, who take full responsibility for the content of the manuscript.

REFERENCES

- [1] H. M. Rohini and S. Prabhavathi, "Automated Poultry Health Monitoring through Acoustic Analysis Using Convolutional Neural Networks," *Engineering, Technology & Applied Science Research*, vol. 15, no. 5, pp. 26339–26343, Oct. 2025, <https://doi.org/10.48084/etasr.11622>.
- [2] Z. Wu, H. Zhang, and C. Fang, "Research on machine vision online monitoring system for egg production and quality in cage environment,"

- Poultry Science*, vol. 104, no. 1, Jan. 2025, Art. no. 104552, <https://doi.org/10.1016/j.psj.2024.104552>.
- [3] M. H. Nidhi, K. Liu, and K. J. Flay, "Review: Multiobject tracking in livestock – from farm animal management to state-of-the-art methods," *animal*, vol. 19, no. 5, May 2025, Art. no. 101503, <https://doi.org/10.1016/j.animal.2025.101503>.
- [4] N. S. N. Abd Aziz, S. Mohd Daud, R. A. Dziyauddin, M. Z. Adam, and A. Azizan, "A Review on Computer Vision Technology for Monitoring Poultry Farm—Application, Hardware, and Software," *IEEE Access*, vol. 9, pp. 12431–12445, 2021, <https://doi.org/10.1109/ACCESS.2020.3047818>.
- [5] F. Pan *et al.*, "Occlusion-Aware Caged Chicken Detection Based on Multi-Scale Edge Information Extractor and Context Fusion," *Animals*, vol. 15, no. 18, Sept. 2025, <https://doi.org/10.3390/ani15182669>.
- [6] V. Shwetha, B. S. Maddodi, V. Laxmi, A. Kumar, and S. Shrivastava, "Latest Trend and Challenges in Machine Learning– and Deep Learning–Based Computational Techniques in Poultry Health and Disease Management: A Review," *Journal of Computer Networks and Communications*, vol. 2024, no. 1, 2024, Art. no. 8674250, <https://doi.org/10.1155/2024/8674250>.
- [7] R. O. Ojo, A. O. Ajayi, H. A. Owolabi, L. O. Oyedele, and L. A. Akanbi, "Internet of Things and Machine Learning techniques in poultry health and welfare management: A systematic literature review," *Computers and Electronics in Agriculture*, vol. 200, Sept. 2022, Art. no. 107266, <https://doi.org/10.1016/j.compag.2022.107266>.
- [8] M. R. Bhuiyan and P. Wree, "Animal Behavior for Chicken Identification and Monitoring the Health Condition Using Computer Vision: A Systematic Review," *IEEE Access*, vol. 11, pp. 126601–126610, 2023, <https://doi.org/10.1109/ACCESS.2023.3331092>.
- [9] D. Kern, T. Schiele, U. Klauck, and W. Ingabire, "Towards Automated Chicken Monitoring: Dataset and Machine Learning Methods for Visual, Noninvasive Reidentification," *Animals*, vol. 15, no. 1, Dec. 2024, <https://doi.org/10.3390/ani15010001>.
- [10] C. Fang, Z. Wu, H. Zheng, J. Yang, C. Ma, and T. Zhang, "MCP: Multi-Chicken Pose Estimation Based on Transfer Learning," *Animals*, vol. 14, no. 12, June 2024, <https://doi.org/10.3390/ani14121774>.
- [11] A. Buslaev, V. I. Iglovikov, E. Khvedchenya, A. Parinov, M. Druzhinin, and A. A. Kalinin, "Albumentations: Fast and Flexible Image Augmentations," *Information*, vol. 11, no. 2, Feb. 2020, <https://doi.org/10.3390/info11020125>.
- [12] Z. Wu, J. Yang, H. Zhang, and C. Fang, "Enhanced Methodology and Experimental Research for Caged Chicken Counting Based on YOLOv8," *Animals*, vol. 15, no. 6, Mar. 2025, <https://doi.org/10.3390/ani15060853>.
- [13] M. Paiano, S. Martina, C. Giannelli, and F. Caruso, "Transfer learning with generative models for object detection on limited datasets," *Machine Learning: Science and Technology*, vol. 5, no. 3, Dec. 2024, Art. no. 035041, <https://doi.org/10.1088/2632-2153/ad65b5>.
- [14] D. Lewy and J. Mańdziuk, "An overview of mixing augmentation methods and augmentation strategies," *Artificial Intelligence Review*, vol. 56, no. 3, pp. 2111–2169, Mar. 2023, <https://doi.org/10.1007/s10462-022-10227-z>.
- [15] M. Trigka and E. Dritsas, "A Comprehensive Survey of Machine Learning Techniques and Models for Object Detection," *Sensors*, vol. 25, no. 1, Jan. 2025, <https://doi.org/10.3390/s25010214>.
- [16] M. Shafiq and Z. Gu, "Deep Residual Learning for Image Recognition: A Survey," *Applied Sciences*, vol. 12, no. 18, Sept. 2022, <https://doi.org/10.3390/app12188972>.
- [17] Y. Hui, J. Wang, and B. Li, "DSAA-YOLO: UAV remote sensing small target recognition algorithm for YOLOv7 based on dense residual super-resolution and anchor frame adaptive regression strategy," *Journal of King Saud University - Computer and Information Sciences*, vol. 36, no. 1, Jan. 2024, Art. no. 101863, <https://doi.org/10.1016/j.jksuci.2023.101863>.
- [18] 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>.
- [19] D. Hendrycks, N. Mu, E. D. Cubuk, B. Zoph, J. Gilmer, and B. Lakshminarayanan, "AugMix: A Simple Data Processing Method to Improve Robustness and Uncertainty." arXiv, Feb. 17, 2020, <https://doi.org/10.48550/arXiv.1912.02781>.
- [20] C. Yin, Y. Xu, S. Zhang, J. Jin, and P. Zhang, "Online image augmentation via regional cross-attention," *Computers and Electrical Engineering*, vol. 120, Dec. 2024, Art. no. 109571, <https://doi.org/10.1016/j.compeleceng.2024.109571>.
- [21] A. M. Buna, "Battery Cage Layer Chicken Detection Dataset." Kaggle, [Online]. Available: <https://www.kaggle.com/datasets/abdulmalikbuna/laying-hens>.