

A lightweight visual system for determining the gender of poultry based on artificial intelligence

Jian Luo

Maide Medical Industrial Equipment Co., Ltd

Yongzhan Luo

Maide Medical Industrial Equipment Co., Ltd

Dinghui Luo

Taizhou University

Yi Gao

Taizhou University

Liangliang Lou

lou0908@126.com

Taizhou University

Article

Keywords: Gender identification of chicks, Artificial intelligence, Basicblock, Object detection, Dynamic and efficient fusion module, Dual Attention, Multi-scale fusion

Posted Date: September 21st, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9409056/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

A lightweight visual system for determining the gender of poultry based on artificial intelligence

Jian Luo^{1,2}, Yongzhan Luo^{1,2}, Dinghui Luo², Yi Gao², Liangliang Lou^{2,3*}

¹Maide Medical Industrial Equipment Co., Ltd., Taizhou 318000, China,²Institute of Intelligent Information Processing, Taizhou University, Taizhou 318000, China,³School of Automation, Southeast University, Nanjing 211189, China;

In recent years, gender determination has become increasingly important in modern poultry production. Although both specialized instrument detection and manual visual identification can be used for chick gender determination, the former is costly and time-consuming, while the latter suffers from a high misclassification rate, limiting their applicability in large-scale modern farms. To address these issues, this study proposes an artificial intelligence-based visual gender identification system for chicks, aiming to deliver a low-cost, efficient, and accurate solution for the industry. In this work, the industry's first chick gonad image dataset is established, and the BasicBlock and Dual Attention Mechanism are integrated to construct the BDF (Basic Dual Attention Fusion Net) classification model. Furthermore, to address feather occlusion and illumination interference problems commonly encountered in real-world scenarios, YBDF is proposed as a YOLO-based object detection-driven chick gender identification system. Experimental results demonstrate that the proposed system achieves an average recognition accuracy of 91.33%. The dataset used in this study is publicly available at: https://github.com/Keypants/Chick_dataset.git.

Corresponding author:

Liangliang Lou

Email Address:

lou0908@126.com

Keywords:

Gender identification of chicks; Artificial intelligence; Basicblock; Object detection; Dynamic and efficient fusion module; Dual Attention; Multi-scale fusion.

1. Introduction

In modern poultry production, male chicks require high-energy diets to promote weight gain, while female chicks need high-protein diets to optimize egg-laying performance. Therefore, early identification of chick gender is crucial for optimizing feed allocation, improving animal welfare, and enhancing economic and environmental sustainability[1,2]. Early gender identification enables gender-specific precision management, which not only effectively prevents feed resource waste but also significantly improves the overall operational efficiency of breeding farms [3].

Traditional chick sex identification methods have generally been categorized into pre-hatch and post-hatch approaches. Pre-hatch methods include embryonic fluid hormone analysis, in which small volumes of allantoic or amniotic fluid are extracted and examined through immunoassays; egg spectral analysis, which detects sex-related differences in light absorption or reflectance patterns; and in-ovo genetic testing, which retrieves a minute amount of embryonic material through micro-perforation of the shell for chromosomal examination [4]. Although these techniques can achieve high accuracy, they depend on costly instruments, invasive sampling, and lengthy laboratory procedures, making them difficult to implement at a large commercial scale [5]. Post-hatch methods, such as feather identification and cloacal examination, rely heavily on visual cues and technician expertise [6]. Feather sexing is performed by observing the relative growth of primary and covert feathers shortly after hatching, while cloacal inspection involves manually exposing the cloacal region to distinguish sex-specific anatomical structures. Despite being rapid, these methods often exhibit accuracy below 80 percent, which can decline sharply due to operator fatigue or breed variability[7]. Overall, both pre-hatch and post-hatch techniques face limitations, including high costs, invasiveness, dependence on human expertise, and poor scalability[8]. These shortcomings underscore the urgent need for more accurate, automated, and economically feasible visual sexing technologies.

The development of artificial intelligence technology has opened new pathways for chick sexing, with recognition technologies based on acoustic and feather image texture features providing new directions for low-cost, high-efficiency identification. Cuan et al. [9] proposed an acoustic feature recognition method that collects chick sound signals and employs compact convolutional neural networks for signal processing and sex classification. Experiments demonstrated that this method achieved 91.25% accuracy in controlled laboratory environments, but identification accuracy would significantly decline in real agricultural scenarios due to environmental noise interference, limiting its application in commercial hatcheries. In contrast, visual recognition technology based on feather features is not affected by environmental noise factors. Soliman-Cuevas et al. [10] integrated VggNet architecture with computer vision technology, performing gender classification through feather feature analysis of chick video sequences, achieving 83.33% accuracy. However, this approach requires extensive manual preprocessing and is computationally intensive, constraining its deployment in large-scale poultry farms.

To address the limitations of acoustic recognition's susceptibility to environmental interference and the insufficient precision of feather feature approaches, this study builds upon the cloacal examination method, which enables direct observation of chick gender characteristics, and constructs a gender identification system based on chick gonad images. Additionally, 1,203 high-resolution one-day-old chick gonad images were collected and annotated, establishing the first standardized chick gonad dataset in China (611 male specimens, 592 female specimens), addressing the data shortage in related research. Compared to feather feature recognition, cloacal examination obtains the most reliable physiological information through direct observation of gonadal organs. In terms of chick gonad morphology, males exhibit distinctive circular protrusions, while females appear relatively flat [11].

However, since chick gonads are concealed beneath down feathers, operators must apply pressure to the tissues surrounding the cloaca to expose the gonads for imaging [13], making fingers and down feathers the primary interference factors in image recognition. To address this challenge, this study proposes YBDF, a YOLO object detection-driven chick gender identification system. Through precise gonad region localization using YOLO [12] and elimination of feather and fecal interference, the system integrates Basicblock [13], and Dual attention mechanisms[14] to construct the BDF Classifier model, thereby achieving reliable chick gender identification.

The organizational structure of this study is as follows: Section 2 elucidates the core technical architecture of the YBDF system, providing detailed descriptions of the design principles and implementation mechanisms of the object detection module and BDF classifier; Section 3 validates system performance through comparative experiments, including object detection module performance evaluation, ablation study analysis, and classifier comparison research, comprehensively assessing the effectiveness of the proposed method from perspectives of accuracy and computational efficiency.

2. Methodology

The framework of the YBDF chick gender detection system is illustrated in Figure 1. The system first employs a YOLO object detection module to locate and preprocess the chick gonad regions, then feeds the processed images into the BDF classifier for gender determination.

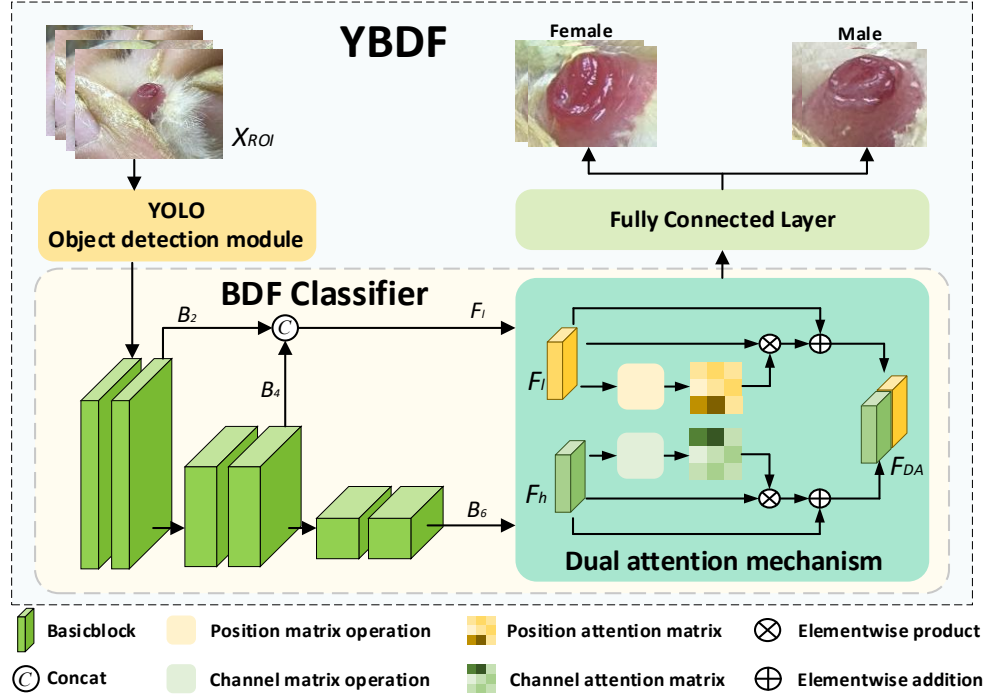


Figure 1. The specific architecture of YBDF

2.1 Dataset preparation and processing

In this study, a dataset of gonadal images from one-day-old yellow broiler chicks was constructed to support the training and evaluation of the proposed YBDF system. A USB 2.0 industrial camera module with a resolution of 1920×1080 pixels was used to capture images of the chick cloacal region, in which gonadal sex-related morphological features could be observed. During image acquisition, each chick was manually positioned, and the cloacal region was gently exposed only when necessary to allow clear visualization of gonadal sex-related features. All procedures were performed carefully to minimize stress and discomfort to the chicks. To ensure uniform illumination and minimize shadows, a cold light source with a color temperature of approximately 5500 K was employed. All images were collected under consistent environmental conditions and saved in JPEG format. After image acquisition, each gonadal image was independently annotated by two experienced poultry specialists according to established gonadal sex-related morphological criteria. Any inconsistent annotations were resolved through discussion between the specialists. A total of 1,203 valid samples were obtained, consisting of 611 male and 592 female chicks. For experimental evaluation, the dataset was randomly divided into training, validation, and test sets at a ratio of 70%, 20%, and 10%, respectively. This dataset provided the foundational data needed to develop and validate the YBDF chick sex classification model.

2.2. Object detection module

Given an input image $I \in R^{H_0 \times W_0 \times 3}$ (H_0 and W_0 denote the height and width of the original input image), YOLO object detection module identifies one (or more) bounding boxes $\hat{b}$ corresponding to the chick gonads. For simplicity, this study assumes that a single main bounding box $\hat{b}$ is chosen (e.g. by selecting the highest objectness score). The bounding box is typically represented by [15]:

$$\hat{b} = YOLO(I) = [\hat{x}, \hat{y}, \hat{w}, \hat{h}] \quad (1)$$

where $\hat{x}$, $\hat{y}$ represents the predicted central coordinates of the chick gonad, $\hat{w}$, $\hat{h}$ represents the predicted width and height of the chick gonad.

Once the bounding box $\hat{b}$ is obtained, YOLO object detection module extracts the corresponding region of interest (ROI) from the original image I . Define a function $\phi(\cdot, \cdot)$ that crops and resizes the ROI to a fixed dimension suitable for the classifier:

$$X_{ROI} = \phi(I, \hat{b}) \in R^{H_1 \times W_1 \times 3} \quad (2)$$

The cropped chick gonad image X_{ROI} serves as the input to the Basicblocks for feature extraction; H_1 and W_1 are the fixed height and width of the ROI after cropping and resizing, ensuring that the input to the Basicblocks is of a consistent spatial dimension.

2.3. BDF classifier

As shown in Figure 1, the BDF classifier serves as the core component of the chick-sexing system, responsible for accurately classifying the gonadal-region images produced by the YOLO object-detection module. By combining deep feature extraction, multi-scale feature fusion, and an optimized attention mechanism, BDF effectively captures and discriminates the distinctive gonadal features of chicks, thereby accomplishing the chick-sexing task.

2.3.1. Feature extraction module

The BasicBlock as shown in Fig. 1, each BasicBlock contains two spatial 3×3 convolutions[16], each followed by batch normalisation and a ReLU activation. A residual shortcut adds the input either directly or through a down-sampling path when the stride equals two. These residual blocks allow the network to learn more complex features while mitigating the vanishing gradient problem[17]. The specific formulation is as follows:

$$\Phi(x) = BN_2(W_2 * \sigma(BN_1(W_1 * x))) \quad (3)$$

$$B_i = \sigma(x_i + \Phi(x_i)), \quad \text{for } i = 1, 2, \dots, 6 \quad (4)$$

where $\Phi(x)$ is the transformation function within the BasicBlock, W_1 and W_2 represent the convolution kernel weight matrices, $*$ denotes the convolution operation, σ represents the ReLU activation function, BN_1 and BN_2 indicate batch normalization operations, and x is the input feature tensor, x_1 is the initial feature obtained from the cropped chick gonad image X_{ROI} , x_i is the input of the i -th Bottleneck, B_i represents the output of the i -th BasicBlock.

2.3.2. Dual attention mechanism

As shown in Figure 1, the Dual attention mechanism consists of position attention [20] and channel attention [21]. By modeling long-range dependencies across two orthogonal dimensions of space and channels, Dual attention mechanism significantly enhances the network's ability to perceive global context and key information. The Dual attention mechanism inputs the low-level features obtained from Feature extraction module, where positional attention first aggregates information from different positions with similar semantics or structures, thereby enhancing the feature representation of key regions. The specific process is as follows:

$$F_l = \text{Concat}(B_2, B_4) \quad (5)$$

$$f = W_f * F_l, \quad g = W_g * F_l, \quad h = W_h * F_l \quad (6)$$

$$s_{ji} = \frac{\exp(f_i \cdot g_j)}{\sum_{i=1}^N \exp(f_i \cdot g_j)} \quad (7)$$

$$E_p = \alpha \sum_{i=1}^N (s_{ji} h_i) + F_l \quad (8)$$

where, F_l is the input of position attention, which is obtained from the low-level features B_2 and B_4 through the $\text{Concat}(\cdot)$ operation; f, g , and h represent the query, key, and value components respectively; The feature transformation matrices W_f, W_g , and W_h are implemented as 1×1 convolution kernels within the positional attention module to generate f, g , and h ; The positional attention matrix S_{ij} captures the attention weights between different spatial locations; The parameter α denotes a learnable residual scaling coefficient for positional attention, initialized to 0; N corresponds to the spatial dimension, calculated as the product of the input feature map's height and width; Following positional aggregation, E_p represents the final output of the positional attention module.

Subsequently, the deep-level features B_6 extracted from the feature extraction module are processed through the channel attention component of the bilateral attention mechanism. This channel attention adaptively modulates the response of individual channels, effectively amplifying maintenance-relevant information while suppressing extraneous interference.

$$a = \text{Conv}(\text{Avgpool}(B_6)) \quad (9)$$

$$c_{ji} = \frac{\exp(a_i \cdot a_j)}{\sum_{i=1}^C \exp(a_i \cdot a_j)} \quad (10)$$

$$E_c = \beta \sum_{i=1}^C (c_{ji} a_i) + a \quad (11)$$

where, a represents the input features for channel attention; c_{ij} denotes the channel attention matrix, indicating the attention scores between different channels; β is the learnable residual scaling coefficient for channel attention, initialized to 0; C represents the number of channels; and E_c denotes the output of channel attention.

Finally, the Dual attention mechanism fuses the outputs from both components, with the result given by:

$$F_{DA} = \text{Conv}_{3 \times 3}(E_p) + \text{Conv}_{3 \times 3}(E_c) \quad (12)$$

where, $\text{Conv}_{3 \times 3}(\cdot)$ denotes a 3×3 convolution, and F_{DA} denotes the output of the bilateral attention module.

2.3.3. Subsubsection

Through Dual attention mechanism, the model progressively integrates and refines multi-scale features. These refined features are then fed into global average pooling to extract channel-dimensional information, subsequently generating probability distributions through fully connected layers with softmax activation function. The classification results are defined as follows:

$$\hat{y} = S(WF(Avg(F_{DA})) + b) \quad (13)$$

$$\hat{y}_{male} + \hat{y}_{female} = 1, \quad \{\hat{y}_{male}, \hat{y}_{female}\} \subset \hat{y} \quad (14)$$

where $\mathcal{F}(\cdot)$ represents the flattening operation; W and b are the weight matrix and bias vector of the final fully connected layer; $S(\cdot)$ denotes the SoftMax activation function that normalizes outputs into probability scores; $\hat{y}$ represents the predicted probability distribution for male and female classifications; $\hat{y}_{male}$ and $\hat{y}_{female}$ are the probabilities for male and female categories respectively, with their sum equal to 1.

Given that chick gender determination constitutes a binary classification task, this study employs cross-entropy loss as the objective function. This function effectively mitigates model overconfidence issues by precisely measuring the discrepancy between predicted probabilities and true labels, thereby effectively suppressing overfitting risks. The definition of this loss function is presented as follows:

$$\mathcal{L} \equiv -\frac{1}{N} \sum_{i=1}^N [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)] \quad (15)$$

where $\mathcal{L}$ represents the total cross-entropy loss for the entire dataset; N is the number of samples in the dataset; y_i is the true label of the i -th sample, which can be 0 or 1; $\hat{y}_i$ is the predicted probability of the model for the i -th sample, ranging from 0 to 1.

3. Results

3.1. Experiment setup

The self-constructed chick gonad image dataset was partitioned into training, testing, and validation sets with a ratio of 70%, 20%, and 10%, respectively. Female samples were labeled as 0, and male samples were labeled as 1. To comprehensively evaluate system performance, this study employed multiple evaluation metrics. For the object detection module, Average Precision (AP) and mean Average Precision (mAP) were used to measure the overlap between predicted and ground-truth bounding boxes. For classification performance, Recall, Accuracy, and F1-score were adopted as core indicators. In addition, the number of parameters and floating-point operations (FLOPs) were calculated to assess computational complexity and efficiency.

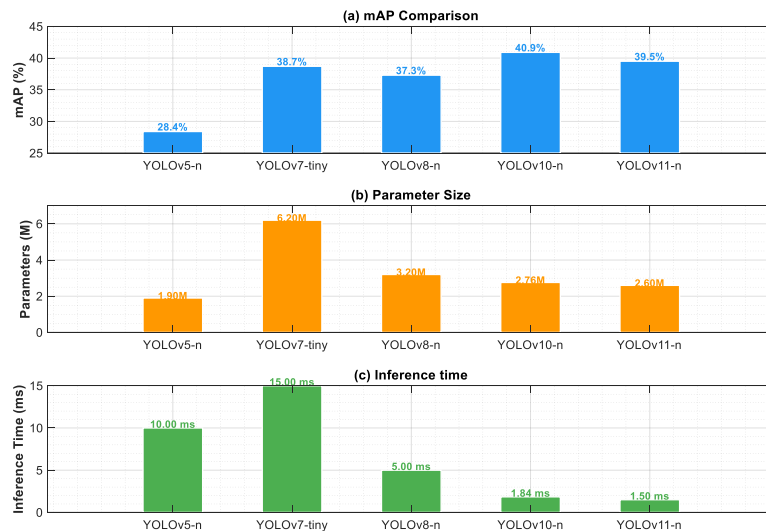
Table 1 Hyperparameters used in the YBDF

Hyperparameter	Value
Image Size	256x256
Batch Size	8
Learning Rate	0.001
Momentum	0.937
Optimizer	SGD
Loss Function	CrossEntropyLoss
weight decay	0.0001
scheduler	StepLR
Number of Epochs	30
Anchor setting	Anchor-free (strides 8, 16, 32)
Flip probability	0.50
Scaling factor	0.80 – 1.20
Color jitter	± 0.20 brightness/contrast/saturation
Mosaic probability	0.50

All experiments were implemented on a Windows operating system using the PyTorch framework. The primary development environment consisted of Python 3.8 and CUDA 11.7, utilizing an NVIDIA GeForce RTX 3090 GPU and an Intel Xeon Silver 4214 CPU running at 2.20 GHz. The entire YBDF model, including both the detection and classification components, was trained strictly following the unified hyperparameter settings summarized in Table 1. These settings included an input image size of 256×256 , a batch size of 8, a learning rate of 0.001 with a momentum of 0.937, an SGD optimizer, a weight decay of 0.0, no learning-rate scheduler, 30 training epochs, and a full set of augmentation and anchor configurations, such as anchor-free detection, a flip probability of 0.50, a scaling factor between 0.80 and 1.20, color jitter of ± 0.20 for brightness, contrast, and saturation, and a Mosaic probability of 0.50.

3.2. Comparison of object detection module

Due to performance differences among various YOLO versions, this study conducts comparative experiments on YOLOv5[18], YOLOv7[19], YOLOv8[20], YOLOv10[21], and YOLOv11[22]. The experiments evaluate object detection module performance across four dimensions: detection accuracy, inference speed, parameter count, and computational complexity, aiming to provide a basis for model selection of the object detection module in chick gender identification tasks.

**Figure 2.** Comparison of object detection models

As shown in Figure 2, in terms of detection accuracy, YOLOv10-n achieves a 40.9% mAP value, surpassing YOLOv11-n with 39.5%, YOLOv8-n with 37.3%, YOLOv7-tiny with 32.4%, and YOLOv5-n with 28.4%, making it the best-performing model among the six. In terms of model complexity, YOLOv10-n has 2.76M parameters, slightly higher than YOLOv5-n and YOLOv11-n, but it reduces the parameter count by 55.5% compared to YOLOv7-tiny, achieving a good balance between optimal accuracy and model lightweighting. Regarding inference efficiency, YOLOv10-n's 1.84ms inference time, although not as fast as YOLOv11-n, shows an 87.7% improvement over YOLOv7-tiny. Considering all three performance metrics, accuracy, model complexity, and inference speed, YOLOv10-n achieves the best balance, making it the ideal choice for the object detection module in this study.

3.3. Ablation experiment

The ablation experiment evaluates the relationship between each module and model accuracy, parameter count, inference time, and YBDF performance, aiming to provide empirical evidence for model optimization strategies and architectural design decisions.

Table 2 Group ablation experiments of YBDF.

Module			Accuracy
YOLOv10	Feature extraction module	Dual attention mechanism	(%)
✓	×	×	55.00 (baseline)
×	✓	×	56.58 (+1.58)
✓	✓	×	90.43 (+35.43)
✓	✓	✓	91.33 (+36.33)

As summarized in Table 2, the baseline YOLOv10 model attains an accuracy of only 55.00%, reflecting the inherent difficulty of classifying cropped gonadal images in the presence of substantial visual noise arising from feather obstruction, finger occlusion, and illumination inconsistencies. Such conditions render shallow features inadequate for capturing the subtle sex-specific morphological cues of chick gonads.

Introducing the Dual Attention mechanism alone results in a minor accuracy increase to 56.58%, indicating that attention without sufficiently expressive feature representations has limited impact. In contrast, the Feature Extraction Module markedly improves accuracy to 90.43%, confirming the necessity of deeper hierarchical features for reliable discrimination. When integrated, the Feature Extraction Module and the Dual Attention mechanism exhibit a synergistic effect, culminating in the full YBDF system achieving an accuracy of 91.33%. This improvement not only demonstrates the complementary nature of the two modules but also highlights the architectural importance of jointly modeling global structure and localized saliency for effective chick sex identification.

3.4. Comparison of classification effect

To evaluate BDF performance, this study selects five representative deep learning architectures for comparison, including widely applied classic CNN[23] architectures (ResNet[24], VggNet[25], and ShuffleNet[26]), as well as ViT architectures that demonstrate advantages in large-scale data processing (DeepViT[27] and ViT[28]). Model performance is assessed across accuracy, F1-score, recall, and computational efficiency metrics.

3.4.1. Loss variation curve

The dynamic changes in training loss and validation loss can intuitively reflect the learning state of the model, providing quantitative reference criteria for classifier performance comparison and model selection. Therefore, this study evaluates the convergence characteristics and generalization capabilities of each classifier on the chick gonad dataset by analyzing loss variations across 30 training epochs.

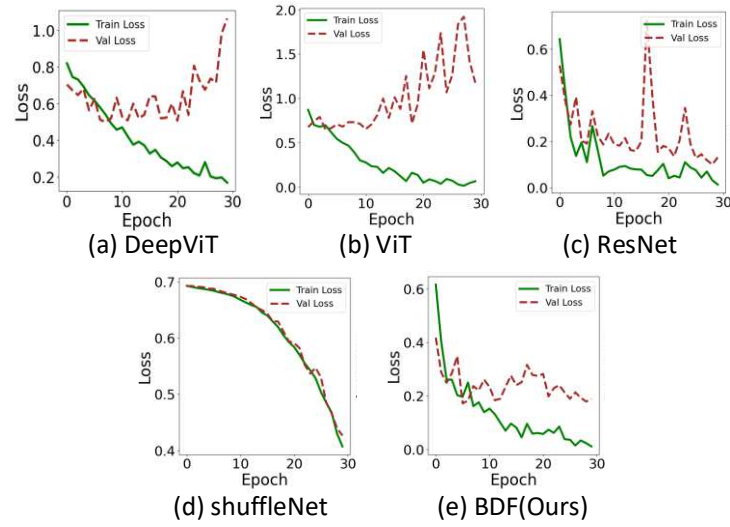


Figure 3. The changes in the training and validation loss.

The training and validation loss curves in Figure 3 clearly illustrate the performance differences between the comparison models and BDF. DeepViT showed a rapid decline in training loss early on, but the validation loss rose to 0.5 in the later stages and stabilized, indicating overfitting. In contrast, BDF maintained synchronized convergence of both training and validation losses to nearly zero, demonstrating excellent generalization ability. ViT exhibited the most severe overfitting, with a convergence trajectory that sharply contrasts with BDF's consistent performance. Although ResNet performed well, with both training and validation losses decreasing to around 0.1, BDF converged faster and achieved a lower final value. ShuffleNet showed steady loss but leveled off at a mid-range, unable to match BDF's performance. Overall, BDF outperformed the baseline models in terms of convergence speed, final accuracy, and generalization.

3.4.2. Analysis of Classification Errors

To intuitively observe the classification performance and potential misclassification patterns of different classifiers, this study adopted t-distributed Stochastic Neighbor Embedding (t-SNE) to visualize the high-dimensional feature representations extracted from the deep learning models. The features used for dimensionality reduction were obtained from the penultimate representation layer of each model, as this layer contains the final discriminative embedding before classification. Specifically, feature vectors were extracted from the last transformer block output of DeepViT and ViT, the global average pooling layer of ResNet, the final stage feature embedding of ShuffleNet, and the high-level feature representation module of the proposed BDF classifier. These feature vectors were then reduced to two dimensions using t-SNE to evaluate the capacity of each model to separate male and female chick samples in the learned feature space.

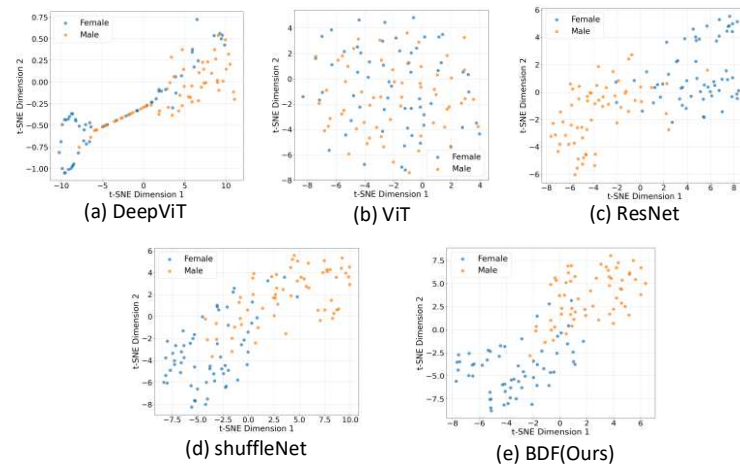


Figure 4. Visualization of t-SNE for different classifiers.

The t-SNE visualization results of the five different classifiers presented in Figure 4 highlight the significant differences between the models and BDF in the chick gender classification task. DeepViT and ViT exhibit considerable overlap between

categories, with a high degree of mixing between red and blue data points, while BDF forms two relatively independent, clearly separated clusters. As a classic convolutional neural network, ResNet shows some improvement in feature separation, but there are still noticeable areas of overlap between categories. In contrast, BDF demonstrates higher intra-class compactness and greater inter-class separation. ShuffleNet shows relatively better clustering performance, with some spatial separation emerging between data points of different genders, but still exhibits a clear gap compared to BDF's distinct class separation. Overall, BDF's t-SNE visualization presents the clearest gender classification boundary and the best clustering performance, confirming its ability to effectively extract key gender features for accurate discrimination and providing a solid technical foundation for practical applications.

3.4.3. Specific Performance evaluation

To comprehensively assess the performance of the proposed BDF classifier, all comparison models were evaluated on the self-constructed test set across key dimensions, including classification accuracy, F1 score, recall, parameter size, and computational cost. The detailed quantitative results are presented in Table 3, where each accuracy value is accompanied by its corresponding 95% confidence interval. These statistical indicators provide a reliable assessment of model stability.

Table 3 Comparison results of different classifiers.

Model	Accuracy	F1-Score (%)	Recall	Params (M)	FLOPs (G)
DeepViT	68.33	70.31	75.00	53.61	3.48
ViT	80.00	79.31	79.31	86.57	17.61
ResNet	89.17	89.43	90.00	11.18	2.38
ShuffleNet	85.00	83.93	78.33	1.26	0.19
BDF(Ours)	91.33	91.53	90.00	7.53	1.82

According to the results in Table 3, the BDF classifier demonstrated the strongest overall performance. It achieved an accuracy of $91.33\% \pm 0.41\%$, an F1 score of 91.53% , and a recall rate of 90.00% . These results clearly surpass the transformer-based models: DeepViT achieved $68.33\% \pm 3.92\%$, and ViT reached $80.00\% \pm 0.48\%$. Compared with ResNet, BDF not only delivered a 2.16 percentage point improvement in accuracy but also reduced the parameter count from 11.18 million to 7.53 million. ShuffleNet remained the most lightweight model with only 1.26 million parameters and 0.19 GFLOPs, but its accuracy of $85.00\% \pm 0.56\%$ was notably lower than that of BDF. Overall, these results show that BDF achieves an effective balance between high accuracy and compact model size, making it highly suitable for deployment in environments with limited computational resources.

3.4.4. Analysis of Model Decision Mechanisms

To complement the feature-distribution analysis provided by t-SNE and gain deeper insight into how different classifiers make decisions, this study further employed Gradient-weighted Class Activation Mapping (Grad-CAM). Unlike t-SNE, which visualizes feature separability, Grad-CAM highlights the specific image regions each model relies on when distinguishing male and female chicks. This allows for a direct assessment of model interpretability and the consistency of each model's decision-making process with anatomical cues.

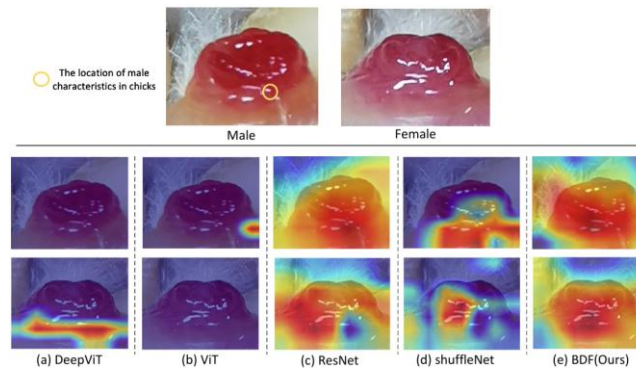


Figure 5. Grad-CAM visualization of decision regions.

The Grad-CAM visualizations in Figure 5 reveal distinct differences in attention patterns. DeepViT and ViT exhibited dispersed and inconsistent activation regions, often focusing on peripheral tissues rather than the central gonadal structures, indicating a weaker ability to localize sex-relevant morphological cues. ResNet showed a more concentrated activation pattern but still

highlighted broader regions surrounding the gonad, suggesting partial but not fully precise localization. ShuffleNet demonstrated improved attention consistency, capturing some key areas with moderate precision, though its activation remained less focused than desirable. In contrast, the proposed BDF classifier produced concentrated activation maps that align closely with the true anatomical regions used by human experts. The BDF consistently focused on the characteristic male protrusion or the flat female morphology, demonstrating strong interpretability and validating its effectiveness in extracting physiologically meaningful features.

3.5. Real-world Deployment Evaluation.

To assess the practical applicability of the proposed YBDF system, real-world validation was conducted in commercial hatchery environments. The system was installed on a modest workstation equipped with an NVIDIA GTX 1050 Ti GPU and an Intel Core i5 processor, representing an affordable hardware configuration commonly found in small and medium-scale production facilities.

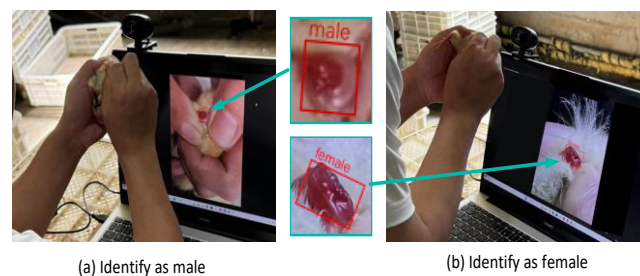


Figure 6. Real-world Deployment of YDF

As shown in Figure 6, the end-to-end processing time per chick, including gonad detection, ROI extraction, and classification, was approximately 85 milliseconds, resulting in a throughput of about 11 to 12 chicks per second. When tested under lighting conditions ranging from 450 to 650 lux and operated by different personnel, the classification accuracy varied by less than 1.5%, indicating strong robustness to environmental changes and operator variability. The software interface shown in Figure 6 enabled real-time visualization and manual verification, and the system operated continuously for six hours without noticeable performance degradation, confirming its feasibility for industrial deployment even on relatively low-cost hardware.

4. Conclusion

The chick-sexing system YBDF, based on gonadal image features, integrates YOLOv10 object detection, Basicbolk for feature extraction, and Dual attention mechanisms, effectively suppressing environmental noise and enhancing the discrimination of subtle gonadal traits. Future work must expand the dataset to cover multiple avian species and environmental conditions, conduct in-field validation across varied farming scenarios, develop hardware solutions with a high cost-benefit ratio, and simultaneously advance economic analysis for large-scale deployment and industry collaboration to guide commercial application.

5. Discussion

The experimental findings demonstrate that the YBDF achieves substantial improvements in chick sex identification accuracy by leveraging precise gonad localization and enhanced multi-scale feature representation. The integration of a YOLO-based detection module with the BDF classifier effectively addresses interference from feathers, manual handling variability, and illumination variation, which frequently undermine the reliability of traditional visual inspection methods. The feature separability and attention localization observed in the analyses indicate that accurate sex prediction depends on both fine structural cues within the gonadal region and broader contextual feature integration, reflecting the importance of combining hierarchical representation learning with attention-driven refinement. Despite these advancements, several methodological challenges require further examination. The current dataset focuses on a single chick breed acquired under controlled imaging conditions, which may limit generalizability to diverse production environments. Future research should include multi-breed datasets, automated image acquisition systems, and domain adaptation techniques to enhance robustness across farming scenarios. These considerations highlight the need for continued refinement before large-scale industrial deployment.

Ethics Statement

The chick inspection and image collection procedures performed in this study involved only brief manual handling consistent with routine hatchery management practices. No invasive procedures, experimental interventions, surgery, drug administration,

euthanasia, or procedures causing pain, distress, or harm were conducted for the purpose of this study. According to the applicable institutional guidelines for animal care and use, these non-invasive husbandry-related procedures were considered exempt from formal ethical review.

Acknowledgement

The present study was funded by Science and Technology Plan Project of Taizhou, China (NO. 24nya19)

Data Availability

The dataset, source code, and trained model weights generated and/or analysed during the current study are publicly available in the GitHub repository: https://github.com/Keypants/Chick_dataset.git.

Informed consent

Written informed consent was obtained from the two experienced poultry specialists involved in chick sex identification and image annotation in this study. The specialists were adults and participated only in professional image annotation and sex identification. No personal data, biological samples, tissue samples, or identifiable information from human participants were collected, analysed, or published in this study.

References

1. Xie, C.; Tang, W.; Yang, C. A review of the recent advances for the in ovo sexing of chicken embryos using optical sensing techniques. *Poultry science* **2023**, *102*, 102906.
2. Romero-Sanchez, H.; Enting, H.; van Eck, L.; van Emous, R.; Neto, F.K.; Leentfaar, E.; Buresh, B. Achieving reproductive performance and quality chicks with modern broiler breeders. *Journal of Applied Poultry Research* **2025**, 100539.
3. Bist, R.B.; Bist, K.; Poudel, S.; Subedi, D.; Yang, X.; Paneru, B.; Mani, S.; Wang, D.; Chai, L. Sustainable poultry farming practices: a critical review of current strategies and future prospects. *Poultry Science* **2024**, *103*, 104295.
4. Bernardi, O.; Ramé, C.; Reverchon, M.; Dupont, J. Expression of chemerin and its receptors in extra-embryonic annexes and role of chemerin and its GPR1 receptor in embryo development in layer and broiler hens. *Poultry Science* **2024**, *103*, 103339.
5. Van Der Hofstadt, M.; l'Helgoualch, N.; Houot-Cernettig, J.; Galindo, T.; Manso, T.; Martins, P.G.A.; Huguenin, J.; Espeut, J.; Molina, F.; Van, T.N.N. Advancing in ovo egg sexing through molecular sex identification of chick embryos using LAMP and RPA assays. *Scientific Reports* **2025**, *15*, 27722.
6. Lv, X.; Sun, C.; Liu, X.; Liu, G.; Gong, W.; Qian, H.; Li, Z.; Wu, J.; Zhu, X.; Song, J. Key events in the process of sex determination and differentiation in early chicken embryos. *Animal Bioscience* **2025**, *38*, 1081.
7. Saetiew, J.; Nongkhunsan, P.; Saenjae, J.; Yodsungnoen, R.; Molee, A.; Jungthawan, S.; Fongkaew, I.; Meemon, P. Automated chick gender determination using optical coherence tomography and deep learning. *Poultry Science* **2025**, *104*, 105033.
8. Ren, Y.; Huang, Y.; Wang, Y.; Zhang, S.; Qu, H.; Ma, J.; Wang, L.; Li, L. A high-performance day-age classification and detection model for chick based on attention encoder and convolutional neural network. *Animals* **2022**, *12*, 2425.
9. Cuan, K.; Li, Z.; Zhang, T.; Qu, H. Gender determination of domestic chicks based on vocalization signals. *Computers and Electronics in Agriculture* **2022**, *199*, 107172.
10. Soliman-Cuevas, H.L.; Linsangan, N.B. Day-Old Chick Sexing Using Convolutional Neural Network (CNN) and Computer Vision. In Proceedings of the 2023 IEEE International Conference on Artificial Intelligence in Engineering and Technology (ICAET), 2023; pp. 45-49.
11. Kannan, A.; Basu, J.; Roy, R.; Pal, M.; Rama Rao, S.; Chatterjee, R.; Ghosh, T.; Ray, H.; Ghosh, A. Gender identification of chicks using vocalisation signals, artificial intelligence and machine learning techniques: Current status and future prospects. *World's Poultry Science Journal* **2025**, *81*, 87-102.
12. Wang, A.; Chen, H.; Liu, L.; Chen, K.; Lin, Z.; Han, J. Yolov10: Real-time end-to-end object detection. *Advances in Neural Information Processing Systems* **2024**, *37*, 107984-108011.

13. Yu, Y.; Huang, J.; Wang, L.; Liang, S. A 1D-inception-ResNet based global detection model for thin-skinned multifruit spectral quantitative analysis. *Food Control* **2025**, *167*, 110823. 359 360
14. Yan, S.; Fan, C.; Zhang, H.; Yang, X.; Tao, J.; Lv, Z. Darnet: Dual attention refinement network with spatiotemporal construction for auditory attention detection. *Advances in Neural Information Processing Systems* **2024**, *37*, 31688-31707. 361 362
15. Tan, L.; Huangfu, T.; Wu, L.; Chen, W. Comparison of RetinaNet, SSD, and YOLO v3 for real-time pill identification. *BMC medical informatics and decision making* **2021**, *21*, 1-11. 363 364
16. Sheng, J.; Zhou, J.; Wang, J.; Ye, P.; Fan, J. Dualmamba: A lightweight spectral-spatial mamba-convolution network for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing* **2024**. 365 366
17. Razavi, M.; Mavaddati, S.; Koohi, H. ResNet deep models and transfer learning technique for classification and quality detection of rice cultivars. *Expert Systems with Applications* **2024**, *247*, 123276. 367 368
18. Gao, W.; Zong, C.; Wang, M.; Zhang, H.; Fang, Y. Intelligent identification of rice leaf disease based on YOLO V5-EFFICIENT. *Crop Protection* **2024**, *183*, 106758. 369 370
19. Zhu, X.; Wang, R.; Shi, W.; Liu, X.; Ren, Y.; Xu, S.; Wang, X. Detection of pine-wilt-disease-affected trees based on improved YOLO v7. *Forests* **2024**, *15*, 691. 371 372
20. Vaghela, R.; Vaishnani, D.; Srinivasu, P.N.; Popat, Y.; Sarda, J.; Woźniak, M.; Ijaz, M.F. Land cover classification for identifying the agriculture fields using versions of yolo v8. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **2025**. 373 374 375
21. Sharma, A.; Kumar, V.; Longchamps, L. Comparative performance of YOLOv8, YOLOv9, YOLOv10, YOLOv11 and Faster R-CNN models for detection of multiple weed species. *Smart Agricultural Technology* **2024**, *9*, 100648. 376 377
22. Khanam, R.; Hussain, M. Yolov11: An overview of the key architectural enhancements. *arXiv preprint arXiv:2410.17725* **2024**. 378
23. Chuquimarca, L.E.; Vintimilla, B.X.; Velastin, S.A. A review of external quality inspection for fruit grading using CNN models. *Artificial Intelligence in Agriculture* **2024**, *14*, 1-20. 379 380
24. Zhou, Y.; Wang, Z.; Zheng, S.; Zhou, L.; Dai, L.; Luo, H.; Zhang, Z.; Sui, M. Optimization of automated garbage recognition model based on resnet-50 and weakly supervised cnn for sustainable urban development. *Alexandria Engineering Journal* **2024**, *108*, 415-427. 381 382 383
25. Rathinam, V.; Sasireka, R.; Valarmathi, K. An Adaptive Fuzzy C-Means segmentation and deep learning model for efficient mammogram classification using VGG-Net. *Biomedical Signal Processing and Control* **2024**, *88*, 105617. 384 385
26. Zhao, H.; Gao, Y.; Deng, W. Defect detection using shuffle Net-CA-SSD lightweight network for turbine blades in IoT. *IEEE Internet of Things Journal* **2024**, *11*, 32804-32812. 386 387
27. Yao, H.; Liu, T.; Li, G.; Qin, C. Screen-shot and Demoiréd image identification based on DenseNet and DeepViT. *Expert Systems with Applications* **2024**, *240*, 122580. 388 389
28. Yao, T.; Li, Y.; Pan, Y.; Mei, T. Hiri-vit: Scaling vision transformer with high resolution inputs. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2024**, *46*, 6431-6442. 390 391