YOLO11-DML: A Lightweight Object Detection Method for Cattle

Zh. Wei[1], X. Zhang[1], X. Li[2], Zh. Yu[1], D. Wei[1*], J. Ni[1]

1- College of Electrical and Information, Northeast Agricultural University, Harbin 150030, China
2- College of Agriculture, Northeast Agricultural University, Harbin 150030, China
(*- Corresponding Author Email: weidonghui@neau.edu.cn)
https://doi.org/10.22067/jam.2026.96943.1454

Abstract

Reliable cow detection in barn environments remains challenging due to visual occlusion, overlapping animals, and cluttered backgrounds. To improve accuracy and real-time performance under such conditions, this study presents YOLO11-DML, an enhanced detection framework built upon YOLO11. To address the decline in recognition accuracy caused by cow overlaps, occlusions, and background interference, this paper has implemented optimisations in three aspects: feature extraction, attention mechanism, and lightweight detection head design: The C3K2-DIMB module is introduced into the Backbone and Neck, enhancing multi-scale feature modelling capabilities through dynamic convolution kernel weights; The mixed local channel attention (MLCA) hybrid local-channel attention mechanism is embedded at the end of the Backbone as an auxiliary feature enhancement module to refine hierarchical feature representations; A lightweight shared convolutional detection head (LSCDH) is designed for the detection head, effectively reducing parameter count and computational overhead while maintaining detection accuracy. Experiments conducted on the CBVD-5 and Dairy Cow datasets show that YOLO11-DML achieves a precision (P) of 92.57%, an F1-score of 89.07%, an mAP@0.5 of 93.11%, an mAP@0.5:0.95 of 59.37%, an inference speed of 105.79 FPS, a parameter count of 2.16M, 5.1 GFLOPs of floating-point operations, and a model size of only 4.5 MB. The research demonstrates that YOLO11-DML achieves high precision while maintaining a lightweight design and real-time performance, providing a feasible solution for multi-object dairy cow detection and behaviour monitoring in smart farms.

Keywords: C3K2-DIMB, Deep learning, Detection, LSCDH, YOLO11-DML

Introduction

Dairy cows are one of the primary production animals in modern animal husbandry, and their health conditions and behavioural traits directly influence the yield and quality of dairy products (Cabrera *et al.*, 2025; Zhu, Liu, Zhang, Liu, & Li, 2026). As China's animal husbandry moves towards intensification and large-scale development, traditional methods of cow behaviour identification that rely on manual observation have gradually revealed issues such as high labour intensity, strong subjectivity, and poor real-time performance, making it difficult to meet the demands for efficient, precise, and automated monitoring in modern smart pastures (Yang, Xu, Zhao, & Song, 2023). Particularly in complex barn environments, factors like varying lighting conditions, occlusions, and cow overlaps pose challenges. Thus, achieving accurate identification of individual cows and their behaviours has become a key research direction in the field of intelligent livestock farming (Bo *et al.*, 2023; Higaki *et al.*, 2025; Wang, Yeh, & Mark Liao, 2024; J. Wang *et al.*, 2024). In recent years, advancements in computer vision and deep learning technologies have provided new approaches for intelligent monitoring in animal husbandry (Qiao *et al.*, 2021; R. Wang *et al.*, 2024).

In the field of intelligent livestock farming, real-time monitoring stands as one of the core metrics for evaluating the performance of detection systems. Therefore, selecting an object detection algorithm capable of adapting to complex barn environments and achieving high detection efficiency holds significant importance for practical deployment. Among numerous object detection networks, the YOLO series models have garnered attention due to their effective balance between speed and accuracy. Siriani *et al.* (2022) improved the YOLOv4 object detection model and combined it with Kalman filtering to achieve

chicken detection and tracking in low-resolution videos. Jiang, Wang, Li, Gouda, and Zhou (2025) proposed a novel real-time monitoring system for the poultry industry based on an attention mechanism-enhanced YOLO-Chicken and BoT-SORT tracking algorithm, achieving an accuracy exceeding that of the original YOLOv8 model by 1.35% and outperforming other object detection algorithms, including Faster R-CNN and RetinaNet, in terms of both accuracy and efficiency. Z. Li *et al.* (2025) proposed integrating the SimAM attention module, residual networks, and a multi-dimensional collaborative attention mechanism (MCA) to form the MCA-SimAM-Resnet (MRS-ATT) module and improve the YOLOv5 model using the LMPDIoU loss function to enhance its capability for beef cattle detection. X. Zhang, Xuan, Xue, Chen, and Ma (2023) introduced a lightweight sheep face recognition model, LSR-YOLO, which incorporates lightweight modules such as ShuffleNetv2 and Ghost, along with the CA attention mechanism, into YOLOv5s, achieving an mAP@0.5 recognition accuracy of 97.8% with a model size of only 9.5 MB.

Currently, scholars have been making successive attempts to improve object detection models from directions such as attention mechanisms and network structure optimisation (B. Li, Fang, & Zhao, 2025; Yang *et al.*, 2023). For instance, introducing multi-scale feature fusion modules to enhance the expressive capacity of feature layers or adopting lightweight convolutional structures to reduce computational overhead. However, in practical applications, there remains a trade-off between improving model accuracy and reducing parameter count, making it challenging to balance detection performance with real-time deployment efficiency.

To address this, this paper proposes an improved cow object detection model, YOLO11-DML, based on the YOLO11 model. Firstly, an improved module named C3K2-DIMB is designed to replace the original C3K2 modules in both the Backbone and Neck, enhancing multi-scale feature modelling capabilities through a dynamic convolution kernel weighting mechanism. Secondly, the MLCA (Mixed Local Channel Attention) mechanism is introduced to strengthen the model's ability to represent complex backgrounds and occluded targets. Finally, a Lightweight Shared Convolution Detection Head (LCSDH) is constructed to effectively reduce the number of parameters and computational load while maintaining detection accuracy. Experimental results on the publicly available datasets CBVD-5 and Dairy Cow demonstrate that the proposed YOLO11-DML model can significantly reduce model complexity while maintaining high detection accuracy, providing an efficient and feasible solution for automated monitoring and health assessment of cow behaviour in smart pastures.

Materials and Methods

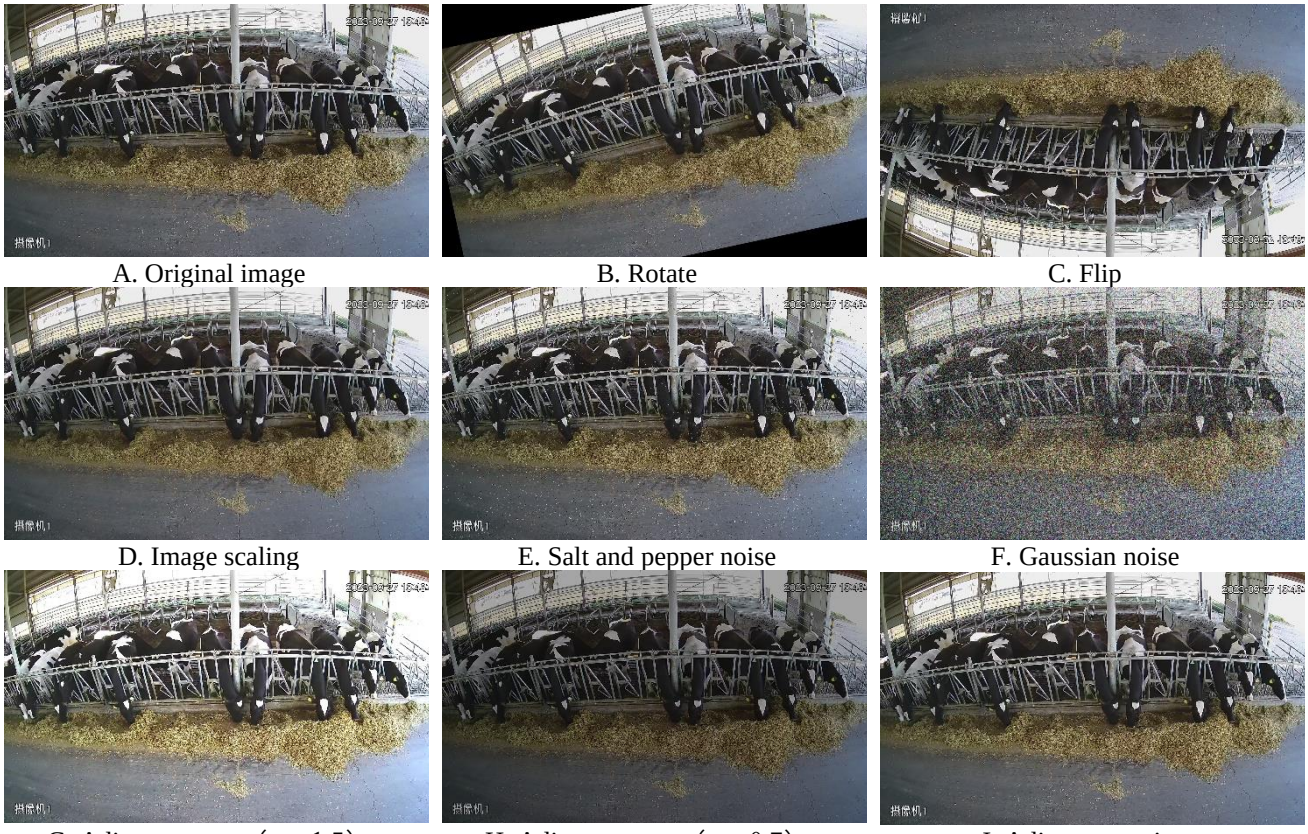
Datasets

In this study, the CBVD-5 (Cow Behaviour Video Dataset-5) was primarily adopted as the dataset. This dataset was constructed by K. Li, Fan, Wu, and Zhao (2024). Through rigorous quality screening and cleaning of the original materials obtained, we first removed samples from the images provided in the dataset that contained annotation errors, missing annotations, or severely occluded targets, ultimately obtaining 3,432 effectively annotated images.

To further enhance the model's robustness under low-light conditions, this study selectively extracted 273 images of cows in night-time scenes from the Dairy Cow dataset available on the Kaggle platform (Twisdu, 2021). This dataset complements CBVD-5, significantly improving the diversity of training samples. After combining the two datasets, a total of 3,705 image samples were obtained, which were subsequently expanded to 10,322 through data augmentation, providing a sufficient data foundation for the training and validation of the deep learning model.

Data Processing

Since deep learning models impose high demands on the quantity and diversity of samples during training, relying solely on a single set of original data is often insufficient to support effective model training. Therefore, when constructing the experimental dataset, systematic data augmentation and pre-processing were carried out on the original samples to enhance the model's robustness and generalisation capabilities. This study employed a variety of data augmentation techniques, such as random rotation, scaling, brightness adjustment, contrast adjustment, and the addition of Gaussian noise and salt-and-pepper noise. Ultimately, the augmented images and their corresponding labels were expanded to a total of 10,322. Data augmentation not only enables dataset expansion but also improves the model's adaptability to different scenarios. The augmented dataset was randomly divided according to a specified ratio, with 60% allocated for model training, 20% for testing, and 20% for validation. The effects of some data augmentation methods are illustrated in Figure 1.



A. Original image B. Rotate C. Flip

D. Image scaling E. Salt and pepper noise F. Gaussian noise

G. Adjust contrast（α = 1.5） H. Adjust contrast（α = 0.7） I. Adjust saturation

Fig. 1. Illustration of the data augmentation effect

YOLO11-DML Network Architecture

In this study, YOLO11 (Khanam & Hussain, 2024) was selected as the baseline model and further improved to meet the specific requirements of cow object detection. YOLO11 features a lightweight and modular network architecture consisting of a Backbone, Neck, and Head, which facilitates the extraction and fusion of multi-scale features. Compared to its predecessor, YOLOv8 (Qiu, 2023), YOLO11 introduces the C3K2 module in the backbone network and adds a C2PSA module after the SPPF layer, enhancing multi-scale feature representation through cross-stage connections and spatial attention mechanisms. These design characteristics enable YOLO11 to maintain high detection accuracy in complex pasture scenarios while

balancing model lightweightness and high inference efficiency, meeting the demands of real-time video surveillance.

In response to the challenges of similar cow appearances and frequent occlusions in pasture video surveillance scenarios, this paper optimises the original YOLO11 network from three aspects. Firstly, dynamic convolution kernel weights are introduced into the backbone network to enhance multi-scale feature representation capabilities. Secondly, a hybrid attention mechanism is fused at the high-level feature stage to strengthen the model's focus on key regions. Finally, a lightweight shared convolution structure is adopted in the detection head to reduce the model's parameter count and computational complexity.

Overall, YOLO11-DML achieves coordinated optimisation of feature extraction, feature fusion, and target prediction processes by making targeted improvements to key functional modules while maintaining the stability of the original network framework. The improved network structure can more accurately characterise cow targets in complex backgrounds and occlusion environments, while balancing detection accuracy and inference efficiency. The overall structure of the model is shown in Figure 2, and the specific designs of each improved module will be further introduced in the next Section.

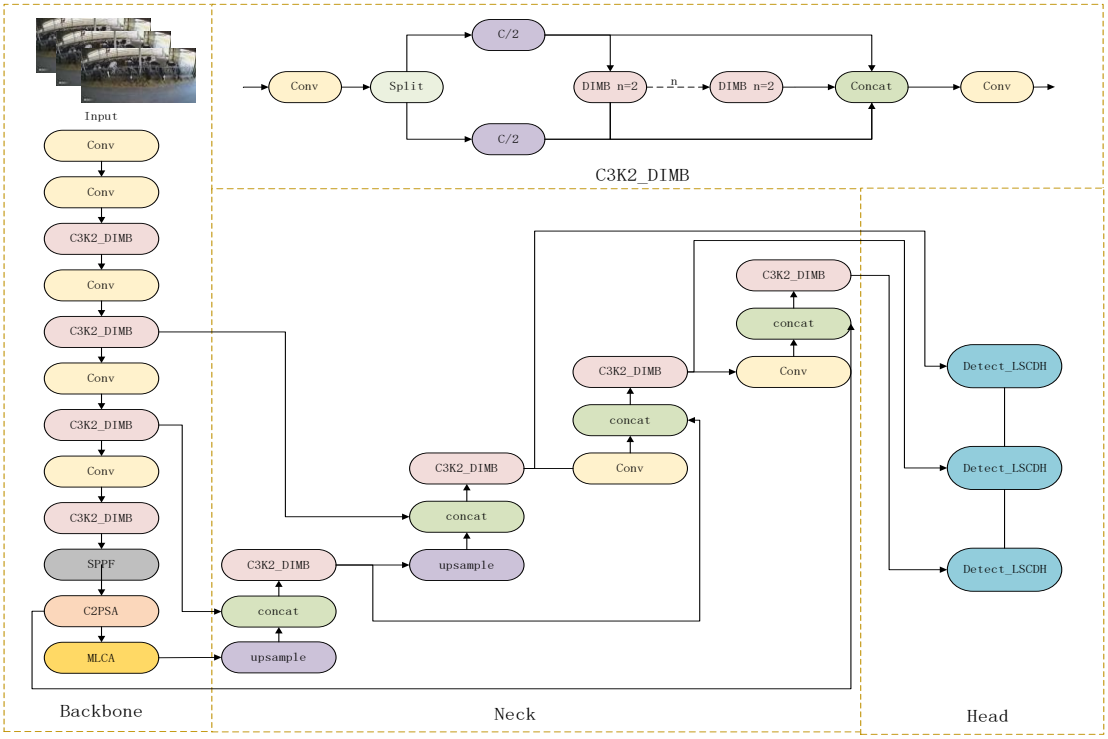


Fig. 2. Overview of the YOLO11-DML network architecture

Model Improvements

C3K2-DIMB Module

In barn surveillance scenarios, cow detection requires the model to capture not only fine-grained details, such as hair texture and body contours, but also global semantic information under conditions involving occlusion, lighting changes, and background interference. The original C3K2 module in YOLO11 uses fixed convolutional kernels, which provide good extraction speed and reasonable accuracy but lack flexibility in adapting to multi-scale variations. As a result, the module struggles to represent both local and global features effectively in complex scenes.

In complex pasture surveillance scenarios, the task of cow detection requires not only

capturing local detailed features (such as fur texture and colour) but also acquiring global semantic information under conditions of occlusion, lighting variations, and background interference. The C3K2 module in traditional YOLO11 employs a fixed convolutional kernel structure, which, while performing well in terms of feature extraction speed and accuracy, still exhibits certain limitations in multi-scale feature adaptability and convolutional operation flexibility. To address these issues, this paper improves the C3K2 modules in the backbone and neck networks by integrating the Inception depthwise convolution from InceptionNeXt (Yu, Zhou, Yan, & Wang, 2024) and the CGLU module from TransNext (Shi, 2024).

Assuming the input feature map X has dimensions C×H×W, the feature map X is processed through parallel K×K square convolutions, M×1 horizontal convolutions, and 1×M vertical convolutions to obtain three output feature maps P2, P3, and P4, each with dimensions C×H×W, where M=3k+2. Simultaneously, adaptive average pooling (AdaptiveAvgPool2d) and 1×1 convolutions are used to generate weight vectors $\alpha 1$, $\alpha 2$, and $\alpha 3$, which are then normalised using Softmax as follow:

$$Y = \sum_{i=1}^{3} \alpha_i * P_i \quad (1)$$

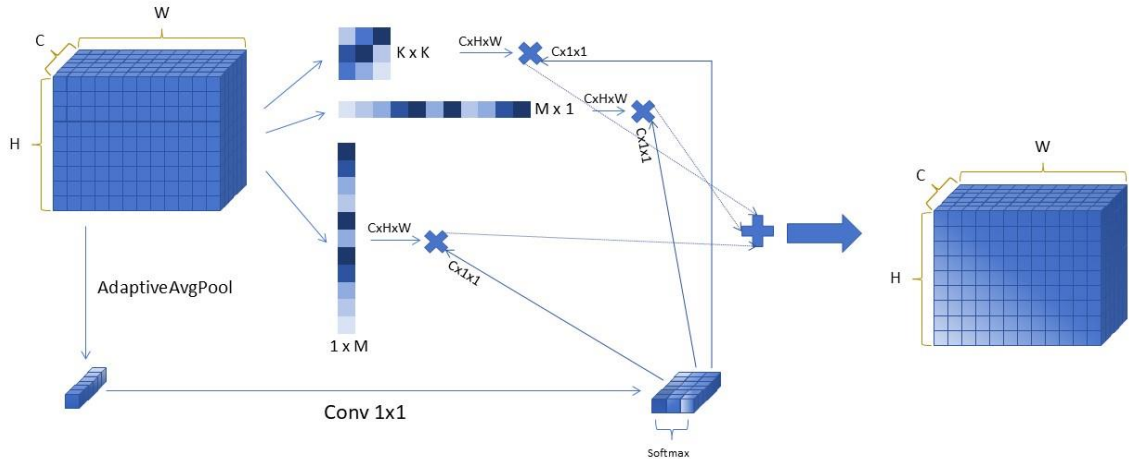


Fig. 3. Structure of the DIDWConv2d module

In this way, during the forward propagation process, the network can automatically select and enhance the most discriminative convolutional branches based on the scale, orientation, and environmental complexity of the cow targets in the image, while simultaneously suppressing irrelevant or redundant feature channels. Ultimately, by integrating the aforementioned improvements, the Dynamic IncMixer Block (DIMB) module is designed as the core feature processing unit of the C3K2-DIMB.

MLCA Attention Mechanism

In real-world pasture surveillance scenarios, the detection of cows is frequently influenced by various factors, such as lighting variations, occlusions, cluttered backgrounds, and differences in camera perspectives. These factors can lead to significant disparities in the saliency of target features across different scales and channels, thereby affecting the model's detection accuracy. Attention mechanisms have been widely adopted in numerous deep learning domains in recent years. Their core idea is to selectively emphasise important features and suppress irrelevant or redundant information under limited computational resources, enabling neural networks to rapidly focus on task-

relevant regions. In the design of lightweight neural networks, common attention modules include Channel Attention (CA) (Hou, Zhou, & Feng, 2021) and Spatial Attention (SA) (Zhang & Yang, 2021), which typically enhance features at a single scale. However, single-scale feature enhancement tends to overlook the complementarity between features at different levels, resulting in the underutilisation of some critical information.

To enhance the model's feature extraction capability in complex scenarios, this paper introduces the Mixed Local Channel Attention (MLCA) mechanism (Wan *et al.*, 2023). Building upon a single attention mechanism, MLCA further integrates channel information, spatial information, local channel information, and global channel information, performing cross-attention computations across multiple scales to strengthen the interaction and fusion of multi-level features. The structure of the MLCA attention mechanism is illustrated in the accompanying Figure 4. This multi-dimensional, multi-scale fused attention mechanism can effectively improve the model's robustness and generalisation ability under varying lighting conditions and occlusion scenarios without significantly increasing computational complexity, thereby enhancing the accuracy and stability of cow detection tasks. In this paper, MLCA primarily serves as an auxiliary feature enhancement module, working in synergy with other improved modules to boost overall performance.

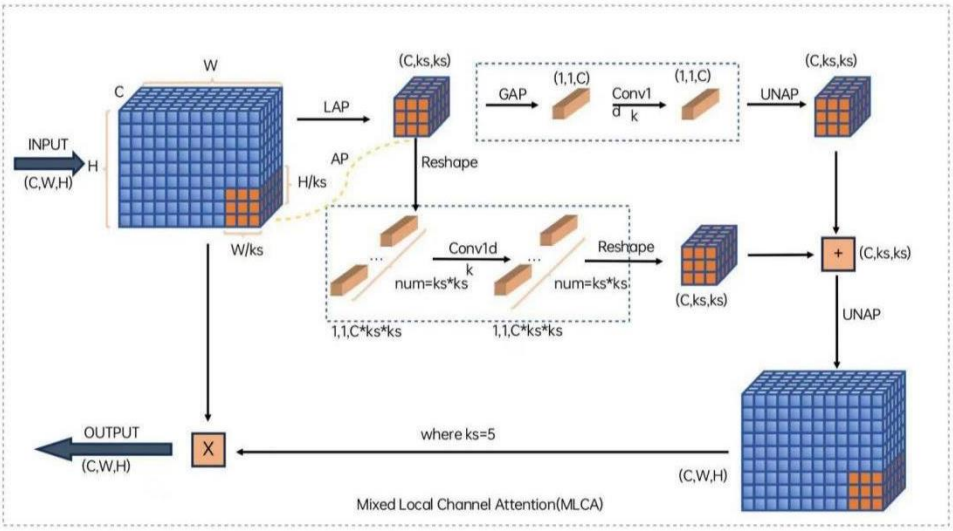


Fig. 4. MLCA attention mechanism

LSCDH Detection Head

YOLO11 inherits the decoupled head design from YOLOv8, utilising independent detection heads at different network levels. This task-decoupling approach effectively reduces mutual interference in multi-category detection, thereby enhancing detection accuracy. The design is implemented through a series of two 3×3 convolutions followed by a 1×1 convolution, but this leads to a significant increase in both the number of parameters and computational complexity.

To address this issue, this paper proposes a Lightweight Shared Convolution Detection Head (LSCDH), as illustrated in the accompanying Figure 5. Initially, multi-scale feature maps (P3, P4, P5) from the feature pyramid undergo a 1×1 convolution followed by Group Normalisation (Conv_GN 1×1). Subsequently, these features are fed into two cascaded 3×3 Conv_GN layers, where different scale branches share the same convolutional weights, substantially reducing the number of parameters and making the model more lightweight. Finally, after processing through the shared convolutions,

the feature maps are separately directed to the regression branch and the classification branch. While utilising shared convolutions, a Scale layer is employed to scale the features, accommodating the detection requirements of features with different resolutions.

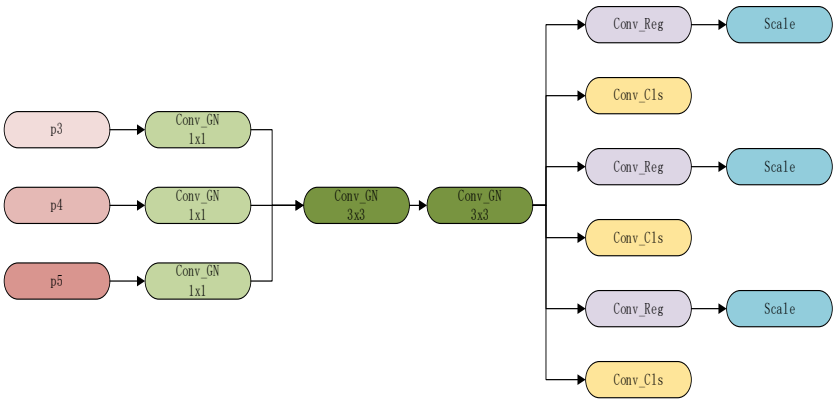


Fig. 5. LSCDH lightweight detection head

Experimental Results

Experimental Setup

For constructing the cattle target recognition model, our experimental platform is configured with an Intel(R) Core i5-9400 CPU (2.9 GHz base frequency, up to 4.1 GHz turbo frequency) and an NVIDIA GeForce GTX 1050 Ti GPU, ensuring robust computational power and acceleration for model training. The software environment utilises PyCharm as the development platform with Python 3.9, based on the PyTorch 2.0.1 deep learning framework, running on a 64-bit Windows 10 OS for stability and compatibility. During experiments, detection images are resized to 640×640 pixels, and the model parameters of the improved LDM-YOLO11 network are optimised using Stochastic Gradient Descent (SGD) through multiple iterations until convergence. Specific training parameters are outlined in Table 1.

Table 1- Training hyperparameters of the model

Parameter	Value
Batch size	16
Epoch	150
Learning rate	0.01
Input size	640

Evaluation Metrics

In this experiment, we utilise precision (P), recall (R), F1 score (F1-score), mean average precision (mAP), the number of parameters (Params), floating-point operations (FLOPs), model size (Model Size), mAP50-95(%), and Frames Per Second (FPS) as evaluation metrics for the model.

Results Analysis

Figure 6 shows a visual comparison of detection results between the baseline YOLO11n model and the proposed YOLO11-DML model.

YOLO11 YOLO11-DML

Fig. 6. Comparison of detection results before and after model improvements

As shown in Figure 6, the original YOLO11 model still shows some missed detections in scenarios where cows overlap or are blocked by pen fences. In contrast, the improved YOLO11-DML model detects target individuals more completely in these complex scenarios. Its object localisation results are more stable. The predicted bounding boxes match the real target areas better. The confidence scores are also more concentrated. Overall, its detection results are more consistent. This indicates that the proposed method has good adaptability and practicality in complex livestock pen environments.

Ablation Study and Result Analysis

To validate the effectiveness of the proposed improvements, ablation experiments were conducted on a cattle image dataset captured from surveillance perspectives, with results summarised in Table 2. The proposed YOLO11-DML model achieved a precision (P) of 92.89%, with 2,158,918 parameters, 5.1 GFLOPs computational cost, and a model size of 4.5 MB.

Table 2- Ablation study results

	P (%)	F1 (%)	mAP50 (%)	Params/M	FLOPs/G	Model Size (MB)	mAP50-95 (%)	FPS
YOLO11	91.72	89	93.22	2.58	6.3	5.2	58.55	152.4
DIMB	91.89	89.05	93.25	2.32	5.8	4.9	59.42	101.71
MLCA	91.82	88.93	93.12	2.58	6.3	5.2	60.12	153.42
LSCDH	92.66	89.40	93.31	2.42	5.6	4.9	60.36	163.5
DIMB+MLCA	92.58	89.04	93.79	2.32	5.8	4.9	59.45	100.66
DIMB+LSCDH	92.20	89	93.23	2.16	5.1	4.5	59.13	104.4
MLCA+LSCDH	92.64	89.13	93.06	2.42	5.6	4.9	59.91	163.85
YOLO11-DML	92.59	89.07	93.11	2.16	5.1	4.5	59.37	105.79

As shown in Table 2, among single-module improvements, the LSCDH detection head demonstrated the most significant enhancement in detection accuracy, improving precision (P) and F1-score by 0.94% and 0.40%, respectively, while maintaining stable performance on the mAP50-95 metric. The C3K2-DIMB module slightly improved detection accuracy while effectively reducing model parameters (from 2.58M to 2.32M), though it introduced a minor inference overhead. In contrast, the MLCA module showed limited performance gains when used independently, with slight fluctuations in some metrics, indicating relatively smaller individual contributions. However, when combined with other modules, MLCA complements multi-layer feature refinement and enhances feature representation, demonstrating synergistic effects. Among

combined improvement strategies, C3K2-DIMB+MLCA achieved the highest mAP50 of 93.79%, while C3K2-DIMB+LSCDH delivered optimal lightweight performance (2.16M parameters, 5.1 GFLOPs) without compromising accuracy. The full YOLO11-DML model, integrating all three improvements, achieved balanced performance: precision increased to 92.59%, with stable F1 and mAP50 scores, while reaching 59.37% on mAP50-95. This highlights the collaborative optimisation across modules.

It should be noted that incorporating the C3K2-DIMB module increased computational complexity during feature extraction, reducing inference speed to 105.79 FPS compared to the baseline YOLO11 (152.4 FPS). Nevertheless, this speed remains well above the real-time processing requirements (25–30 FPS) of typical pasture video surveillance systems, demonstrating strong potential for real-world applications while improving detection accuracy and multi-scale feature representation. This reflects a reasonable trade-off between precision and inference efficiency in object detection tasks, confirming that the proposed lightweight design achieves a favourable balance between performance and speed.

Effectiveness Analysis of the Improved C3K2 Module

To validate the effectiveness of the improved C3K2 module in feature extraction and target representation capabilities, this section introduces three enhanced designs: C3K2-DIMB, C3K2-SWC (D. Li, Li, Chen, & Li, 2025), and C3K2-Dblock (Feijoo, Benito, Garcia, & Conde, 2025), followed by ablation experiments under identical conditions. Experimental results are summarised in Table 3.

Table 3- Performance comparison of improved C3K2 variants

	P (%)	R (%)	F1 (%)	mAP50 (%)	mAP50-95 (%)
C3K2-DIMB	91.89	86.38	89.05	93.25	59.42
C3K2-DBlock	90.26	82.30	86.10	90.82	56.23
C3K2-SWC	90.24	83.52	86.75	91.22	55.42

As shown in Table 3, compared to the original C3K2 module in the baseline YOLO11 model, the improved C3K2-DIMB module achieves significant performance gains across all metrics. Specifically, relative to C3K2-Dblock, C3K2-DIMB demonstrates 1.63%, 4.08%, 2.95%, 2.43%, and 3.19% higher precision, recall, F1-score, mean Average Precision (mAP), and mAP50-95, respectively. Similarly, compared to C3K2-SWC, C3K2-DIMB improves these metrics by 1.65%, 2.86%, 2.30%, 2.03%, and 4.00%, respectively.

These results indicate that C3K2-DIMB outperforms the other two variants in both detection accuracy and stability, confirming the efficacy of integrating the DIMB module into the C3K2 architecture.

Comparison with Different Models

To further validate the effectiveness of the proposed YOLO11-DML model in cow detection tasks, comparative experiments were conducted against the original YOLO11 model as well as several mainstream lightweight detectors, including YOLOv8n, YOLOv9**Error! Reference source not found.** (C.-Y. Wang *et al.*, 2024), YOLOv12n (Tian, Ye, & Doermann, 2025) and Hyper-YOLO (Feng *et al.*, 2024).

As shown in Table 4, YOLO11-DML demonstrates a favourable trade-off between detection accuracy and model complexity. It achieves an mAP@0.5 of 93.11% and a precision of 92.59%, while maintaining lower parameter counts and computational overhead. Compared to the original YOLO11n and YOLOv8n, YOLO11-DML reduces parameters, FLOPs, and model size while sustaining comparable detection performance, validating the effectiveness of our structural optimisation strategies. Although Hyper-

YOLO slightly outperforms YOLO11-DML in detection accuracy, its significantly higher parameter count and computational complexity hinder deployment in resource-constrained environments. In contrast, YOLO11-DML stabilises detection performance with 2.16M parameters, 5.1G FLOPs, and a model size of just 4.5MB, achieving a superior balance between lightweight design and accuracy.

Overall, YOLO11-DML exhibits strong practicality and engineering adaptability in complex cattle barn environments, particularly for real-time applications with strict constraints on model size and computational resources. It provides an effective solution for cattle target detection in embedded systems and intelligent livestock farming platforms.

Table 4- Detection performance of different models

	P (%)	F1 (%)	mAP50 (%)	Params/M	FLOPs/G	Model Size (MB)
YOLO11n	91.72	89	93.22	2.58	6.3	5.2
Yolov8n	91.89	89.30	93.30	2.68	6.8	5.4
Yolov9t	91.91	88.73	93.04	1.73	6.4	4.0
YOLOv12n	90.72	89.25	93.08	2.51	5.8	5.2
Hyper-YOLO	92.89	90.26	93.99	3.62	9.5	7.3
YOLO11-DML	92.59	89.07	93.11	2.16	5.1	4.5

Discussion

Model Performance

To address the decline in detection accuracy caused by factors such as fence occlusions, background interference, and highly similar cattle appearances during multi-object cattle detection, this paper proposes an improved detection model, YOLO11-DML, based on the YOLO11 architecture. The model enhances detection performance in complex livestock farm environments through structural optimisation and attention mechanisms. Specifically, the C3K2-DIMB module strengthens focus on primary cattle features, the MLCA module optimises multi-layer feature interaction and fusion, and the LCSDH module reduces model parameters and computational complexity while maintaining performance.

Experimental results demonstrate that YOLO11-DML achieves a precision (P) of 92.57% and an F1-score of 89.07%, with 2.16M parameters, 5.1 GFLOPs computational cost, a model size of 4.5 MB, and an inference speed of 105.79 FPS. This balance of accuracy and real-time performance meets the demands of target detection in complex cattle barn environments.

Future Work

While YOLO11-DML demonstrates strong performance on existing cattle barn datasets, the experiments were primarily conducted in barns with similar structures and environments, leaving its cross-domain generalisation capability unverified. Future work will focus on testing the model on datasets featuring multi-breed cattle, heterogeneous barn environments, and diverse camera configurations, while exploring the integration of temporal video information to enhance multi-object cattle detection and behaviour recognition performance.

Conclusion

This study proposes the YOLO11-DML model, a cattle object detection model that achieves a balance between detection accuracy and computational efficiency. The model introduces the C3K2-DIMB module on the basis of YOLO11 to enhance the representation of primary cattle features, incorporates the MLCA (Mixed Local-Channel Attention) mechanism at the tail of the backbone to optimise multi-layer feature interactions, and designs a Lightweight Shared Convolutional Detection Head (LCSDH) at the detection head to reduce parameter count and computational complexity. Experimental results show that, on the existing dataset, YOLO11-DML achieves a precision of 92.57% and an F1-score of 89.07%, with a

parameter count of 2.16M, floating-point operations of 5.1 GFLOPs, and a model size of 4.5 MB. These results reflect its ability to balance detection performance and lightweight design in complex farm environments, providing a feasible solution for real-time cattle detection.

Acknowledgements

This work was conducted at the College of Electrical and Information, Northeast Agricultural University, Harbin 150030, China.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Zhihang Wei: Responsible for conceptualisation, methodology design, and model innovation/improvement. He developed the research framework and contributed to the optimization of the proposed models and methods.

Xupeng Zhang: Responsible for conducting the experiments and validation. He carried out the experimental studies, processed the experimental data, and analyzed the results.

Xiaoqian Li: Responsible for visualisation and review and editing services. She prepared the figures, performed data visualization, and contributed to language polishing of the manuscript.

Zhiyu Yu: Responsible for data acquisition and manuscript preparation. He participated in data collection, manuscript writing, and formatting.

Jing Ni: Responsible for supervision and technical advice. He provided guidance on the research methodology, experimental design, and manuscript revision.

Donghui Wei: Responsible for project administration, supervision, and review and editing services. He coordinated the research activities, supervised the overall project, and finalized the manuscript for submission.

References

1. Bo, L., Yuefeng, L., Xiang, B., Yue, W., Haofeng, L., & Xuan, L. (2023). Research on Dairy Cow Identification Methods in Dairy Farm. *Indian Journal of Animal Research, 57*(12). https://doi.org/10.18805/IJAR.BF-1660
2. Cabrera, V. E., Bewley, J., Breunig, M., Breunig, T., Cooley, W., De Vries, A., ..., & Greenfield, R. (2025). Data integration and analytics in the dairy industry: challenges and pathways forward. *Animals, 15*(3), 329. https://doi.org/10.3390/ani15030329
3. Feijoo, D., Benito, J. C., Garcia, A., & Conde, M. V. (2025). Darkir: Robust low-light image restoration. *Proceedings of the Computer Vision and Pattern Recognition Conference*, 10879-10889. https://doi.org/10.48550/arXiv.2412.13443
4. Feng, Y., Huang, J., Du, S., Ying, S., Yong, J.-H., Li, Y., ..., & Gao, Y. (2024). Hyper-yolo: When visual object detection meets hypergraph computation. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 47*(4), 2388-2401. https://doi.org/10.1109/TPAMI.2024.3524377
5. Higaki, S., Menezes, G. L., Ferreira, R. E., Negreiro, A., Cabrera, V. E., & Dórea, J. R. (2025). Objective dairy cow mobility analysis and scoring system using computer vision–based keypoint detection technique from top-view 2-dimensional videos. *Journal of Dairy Science, 108*(4), 3942-3955. https://doi.org/10.3168/jds.2024-25545
6. Hou, Q., Zhou, D., & Feng, J. (2021). Coordinate attention for efficient mobile network design. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 13713-13722. https://doi.org/10.1109/CVPR46437.2021.01350
7. Jiang, D., Wang, H., Li, T., Gouda, M. A., & Zhou, B. (2025). Real-time tracker of chicken for poultry based on attention mechanism-enhanced YOLO-Chicken algorithm. *Computers and*

Electronics in Agriculture, 237, 110640. https://doi.org/10.1016/j.compag.2025.110640
8. Khanam, R., & Hussain, M. (2024). Yolov11: An overview of the key architectural enhancements. *arXiv preprint arXiv:2410.17725*. https://doi.org/10.48550/arXiv.2410.17725
9. Li, B., Fang, J., & Zhao, Y. (2025). RTDETR-Refa: a real-time detection method for multi-breed classification of cattle. *Journal of Real-Time Image Processing, 22*(1), 38. https://doi.org/10.1007/s11554-024-01613-7
10. Li, D., Li, L., Chen, Z., & Li, J. S. (2025). Small Convolutional Kernel with Large Kernel Effect. *Proceedings of the Computer Vision and Pattern Recognition Conference, Nashville, TN, USA*, 10-17. https://doi.org/10.48550/arXiv.2401.12736
11. Li, K., Fan, D., Wu, H., & Zhao, A. (2024). A new dataset for video-based cow behavior recognition. *Scientific Reports, 14*(1), 18702. https://doi.org/10.1038/s41598-024-65953-x
12. Li, Z., Zhang, Y., Kang, X., Mao, T., Li, Y., & Liu, G. (2025). Individual Recognition of a Group Beef Cattle Based on Improved YOLO v5. *Agriculture, 15*(13), 1391. https://doi.org/10.3390/agriculture15131391
13. Qiao, Y., Kong, H., Clark, C., Lomax, S., Su, D., Eiffert, S., & Sukkarieh, S. (2021). Intelligent perception for cattle monitoring: A review for cattle identification, body condition score evaluation, and weight estimation. *Computers and Electronics in Agriculture, 185*, 106143. https://doi.org/10.1016/j.compag.2021.106143
14. Qiu, G. J. a. A. C. a. J. (2023). Ultralytics YOLOv8 (Version 8.0.0). Retrieved from https://github.com/ultralytics/ultralytics
15. Shi, D. (2024). Transnext: Robust foveal visual perception for vision transformers. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 17773-17783. https://doi.org/10.48550/arXiv.2311.17132
16. Siriani, A. L. R., Kodaira, V., Mehdizadeh, S. A., de Alencar Nääs, I., de Moura, D. J., & Pereira, D. F. (2022). Detection and tracking of chickens in low-light images using YOLO network and Kalman filter. *Neural Computing and Applications, 34*(24), 21987-21997. https://doi.org/10.1007/s00521-022-07664-w
17. Tian, Y., Ye, Q., & Doermann, D. (2025). Yolov12: Attention-centric real-time object detectors. *arXiv preprint arXiv:2502.12524*. https://doi.org/10.48550/arXiv.2502.12524
18. Twisdu. (2021). *dairy cow*. Retrieved from: https://www.kaggle.com/datasets/twisdu/dairy-cow
19. Wan, D., Lu, R., Shen, S., Xu, T., Lang, X., & Ren, Z. (2023). Mixed local channel attention for object detection. *Engineering Applications of Artificial Intelligence, 123*, 106442. https://doi.org/10.1016/j.engappai.2023.106442
20. Wang, C.-Y., Yeh, I.-H., & Mark Liao, H.-Y. (2024). Yolov9: Learning what you want to learn using programmable gradient information. *European conference on computer vision*, 1-21. https://doi.org/10.1007/978-3-031-72751-1_1
21. Wang, J., Dai, B., Li, Y., He, Y., Sun, Y., & Shen, W. (2024). An intelligent edge-IoT platform with deep learning for body condition scoring of dairy cow. *IEEE Internet of Things Journal, 11*(10), 17453-17467. https://doi.org/10.48550/arXiv:2402.13616v2
22. Wang, R., Gao, R., Li, Q., Zhao, C., Ru, L., Ding, L., ..., & Ma, W. (2024). An ultra-lightweight method for individual identification of cow-back pattern images in an open image set. *Expert Systems with Applications, 249*, 123529. https://doi.org/10.1016/j.eswa.2024.123529
23. Yang, L., Xu, X., Zhao, J., & Song, H. (2023). Fusion of RetinaFace and improved FaceNet for individual cow identification in natural scenes. *Information Processing in Agriculture*. https://doi.org/10.1016/j.inpa.2023.09.001
24. Yu, W., Zhou, P., Yan, S., & Wang, X. (2024). Inceptionnext: When inception meets convnext. *Proceedings of the IEEE/cvf conference on computer vision and pattern recognition*, 5672-5683. https://doi.org/10.1109/CVPR52733.2024.00542

25. Zhang, Q.-L., & Yang, Y.-B. (2021). Sa-net: Shuffle attention for deep convolutional neural networks. *ICASSP 2021-2021 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, 2235-2239. https://doi.org/10.48550/arXiv:2102.00240
26. Zhang, X., Xuan, C., Xue, J., Chen, B., & Ma, Y. (2023). LSR-YOLO: A high-precision, lightweight model for sheep face recognition on the mobile end. *Animals*, *13*(11), 1824. https://doi.org/10.3390/ani13111824
27. Zhu, M., Liu, Y., Zhang, H., Liu, X., & Li, Z. (2026). Depth-OC-SORT-based multi-object tracking of cattle in real-world farming scenarios. *Computers and Electronics in Agriculture*, *240*, 111151. https://doi.org/10.1016/j.compag.2025.111151

YOLO11-DML: یک روش سبک ساختار (کم‌حجم) تشخیص گاو در گاوداری

ژیهانگ وی[۱]، شوپنگ ژانگ[۱]، شیائوکیان لی[۲]، ژیو یو[۱]، دونگهوی وی[۱*]، جینگ نی[۱]

تاریخ دریافت: ۱۴۰۴/۰۹/۲۱
تاریخ پذیرش: ۱۴۰۴/۱۲/۰۹

چکیده

تشخیص مطمئن گاو در محیط‌های اصطبل به‌دلیل انسداد بصری، همپوشانی دام‌ها و پس‌زمینه‌های شلوغ همچنان چالش برانگیز است. برای بهبود دقت و عملکرد بلادرنگ در چنین شرایطی، این مطالعه YOLO11-DML را ارائه می‌دهد، که یک چارچوب تشخیص بهبودیافته ساخته‌شده بر اساس YOLO11 است. برای رفع کاهش دقت تشخیص ناشی از همپوشانی گاوها و دیگر اجسام و تداخل پس‌زمینه، این مقاله بهینه‌سازی‌هایی را در سه جنبه انجام داده است: استخراج ویژگی، مکانیسم توجه و طراحی تشخیص سبک: ماژول C3K2-DIMB در Backbone و Neck معرفی شده است و قابلیت‌های مدل‌سازی ویژگی چند مقیاسی را از طریق وزن‌های هسته کانولوشن پویا افزایش می‌دهد. مکانیسم توجه کانال محلی ترکیبی (MLCA) در انتهای Backbone به‌عنوان یک ماژول بهبود ویژگی کمکی برای اصلاح نمایش‌های سلسله مراتبی ویژگی تعبیه شده است. یک سر تشخیص کانولوشن مشترک سبک (LSCDH) برای سر تشخیص طراحی شده است که به‌طور موثر تعداد پارامترها و سربار محاسباتی را کاهش می‌دهد و در عین حال دقت تشخیص را حفظ می‌کند. آزمایش‌های انجام‌شده بر روی مجموعه داده‌های CBVD-5 و داده‌های گاو شیری نشان می‌دهد که YOLO11-DML به دقت (P) ۹۲/۵۷ درصد، امتیاز F1 برابر با ۸۹/۰۷ درصد، mAP@0.5 برابر با ۹۳/۱۱٪، mAP@0.5:0.95 برابر با ۵۹٪/۳۷، سرعت استنتاج ۱۰۵/۷۹ FPS، تعداد پارامترهای ۲/۱۶ M، ۵/۱ GFLOP عملیات ممیز شناور و اندازه مدل تنها ۴/۵ مگابایت دست می‌یابد. این تحقیق نشان می‌دهد که YOLO11-DML با حفظ طراحی سبک و عملکرد بلادرنگ، به دقت بالایی دست می‌یابد و یک راه‌حل عملی برای تشخیص چند شیء و نظارت بر رفتار گاوهای شیری در مزارع هوشمند ارائه می‌دهد.

واژه‌های کلیدی: C3K2-DIMB، LSCDH، YOLO11-DML، تشخیص، یادگیری عمیق

۱- دانشکده برق و اطلاعات، دانشگاه کشاورزی شمال شرقی، هاربین ۱۵۰۰۳۰، چین
۲- دانشکده کشاورزی، دانشگاه کشاورزی شمال شرقی، هاربین ۱۵۰۰۳۰، چین
(*- نویسنده مسئول: Email: weidonghui@neau.edu.cn)

https://doi.org/10.22067/jam.2026.96943.1454