Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

ECL-YOLOv8: An improved YOLOv8-based algorithm for detecting cattle faces and muzzles in complex open pasture environments

Juan Liu[1]
Shanxi Agricultural University
Shuai Zhang[1]
Shanxi Agricultural University
Yuhang Li[1]
Shanxi Agricultural University
Hao Zhang[1]
Shanxi Agricultural University
Zishuai Jia[1]
Shanxi Agricultural University
Yuning Ding[1]
Shanxi Agricultural University
Runa Liu[1]
Shanxi Agricultural University
Jie Hu[1]
hujie@sxau.edu.cn

Shanxi Agricultural University

Research Article

Keywords: Cattle Facial and Muzzle Recognition, Deep Learning, YOLOv8, Multiscale Features

Posted Date: November 6th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11465962/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

ECL-YOLOv8: An improved YOLOv8-based algorithm for detecting cattle faces and muzzles in complex open pasture environments

Authors

Juan Liu[1], Shuai Zhang[1], Yuhang Li[1], Hao Zhang[1], Zishuai Jia[1], Yuning Ding[1], Runa Liu[1], Jie Hu[1]*

Affiliations

[1] College of Information Science and Engineering, Shanxi Agricultural University, No. 1 Mingxian South Road, Taigu District, Jinzhong 030801, Shanxi, China

Corresponding author

Correspondence: Jie Hu

Email: hujie@sxau.edu.cn

Abstract

Background

An efficient and reliable method for detecting cattle faces and muzzles supports automated livestock management and intelligent detection.Both the face and muzzle of cattle contain visual information that exhibits significant inter-individual variation. Computer vision-based non-contact identification methods offer advantages such as high efficiency, low cost, and animal welfare friendliness. However, much of the current research is limited to close-up shots or manually selected images of cattle faces or muzzles, with little consideration given to identifying facial or nasal locations in the complex environment of open pastures. This study aims to evaluate the model's robustness in complex environments by conducting multi-angle imaging under various conditions, including varying distances, obstructions from livestock facilities, cattle grazing, muddy faces, and exposure variations. Therefore, this study uses YOLOv8 as the foundation to design the ECL-YOLOv8 model for the simultaneous detection of cattle faces and muzzles in complex open pasture environments.

Results

Experiments conducted on our in-house dataset show that the ECL-YOLOv8 model achieved an accuracy of 92.6%, a recall of 81.7%, and a mean precision of 87.4%. Compared to YOLOv8, ECL-YOLOv8

achieved improvements of 1, 4.2, and 3.1 percentage points in accuracy, recall, and mean precision, respectively. Specifically, the mean precision for muzzle detection was 82.4%, which is 5.3 percentage points higher than the baseline. These results demonstrate that ECL-YOLOv8 exhibits greater robustness under the complex conditions found on ranches.

Conclusions

This study investigates the problem of accurately recognizing cattle faces and muzzles in the complex environment of open pastures and proposes a detection model—ECL-YOLOv8—based on an improved YOLOv8 architecture. The model effectively enhances the detection capabilities of cattle faces and muzzles in complex pasture environments. In particular, the model demonstrates significant advantages in muzzle detection performance and adaptability to complex backgrounds. Future work will focus on improving inference efficiency and model lightweighting to further support practical deployment, thereby better serving real-time monitoring and intelligent operations in actual pastures and providing more reliable technical support for smart agriculture applications.

Keywords Cattle Facial and Muzzle Recognition · Deep Learning · YOLOv8 · Multiscale Features

Background

The livestock industry plays a vital role in the agricultural economy and is a key sector for ensuring national food security and promoting income growth among farmers [1]. As an integral part of modern livestock farming, beef cattle farming relies on precise individual identification as a prerequisite for implementing automated technologies such as intelligent weight measurement, body condition scoring, conformation assessment, and behavioral monitoring; it also serves as the foundation for modern ranch development and precision cattle farming [2,3]. Currently, cattle individual identification technologies are primarily divided into two major categories: contact-based and contactless [4], as shown in Table 1. Compared to traditional biometric methods, contactless classification technologies based on facial and muzzle image recognition offer advantages such as high efficiency, low cost, and animal welfare friendliness, making them an important direction for the development of intelligent beef cattle farming [5]. In light of this, the development of contactless facial and muzzle recognition technologies for beef cattle is imperative.

Table 1：Comparison of Traditional Biometric Methods for Cattle Identification

Methods	Advantages	Disadvantages
Traditional Recognition Systems		
RFID Microchips/Ear Tags: Implanted or attached tags that store a unique ID	Automatic tracking, reliable, and the most widely used	Expensive, with a risk of duplication or loss
Branding: Applying a hot iron to the skin to create a unique mark	Permanent and cost-effective	Pain; risk of duplication
Ear Notch: A specific incision in the ear	Reusable, can incorporate sensors	Suffering, Finite Uniqueness
Neck Tag: An RFID chip attached to a collar	Reusable, can incorporate sensors	May break or be lost
Biometric System		
Nose Print: Using a Unique Nose Print Pattern	Non-invasive, permanent	Close-up imaging is required
Retinal Images: Using Unique Retinal Vein Patterns	Highly Accurate	Disturbing, expensive
Iris Image: Using Unique Iris Patterns	Highly Accurate	Disturbing, expensive
Nasal Ridge Patterns: Algorithmic Analysis of Nasal Ridge Patterns	Contactless, reliable	High-quality imaging is required

Given these limitations, previous studies have found that both the face and the muzzle of cattle contain visual information that exhibits significant inter-individual variation, making them two regions frequently utilized in cattle visual recognition [6]. The face of cattle has a relatively stable outline, eye position, and facial structure, which are sufficient to distinguish between different individuals; the texture of the muzzle is similar to human fingerprints, offering a high degree of uniqueness between individuals and strong discriminatory power, and is widely used in cattle recognition research [7].

In recent years, object detection methods such as Faster R-CNN, YOLO, and DETR have been applied to numerous agricultural computer vision tasks [8,9,10,11]. As for YOLO, its end-to-end design, fast inference speed, and relatively low computational requirements for achieving high accuracy make it a viable option for animal detection [12,13,14]; consequently, numerous studies have discussed and analyzed cattle faces and muzzles as examples.

The muzzle is one of the earliest body parts studied in biometric research and, like fingerprints, has its own distinctive pattern of dots and ridges [15]. Hojaeipour et al. combined YOLO with deep transfer learning to detect cow muzzles and then used ResNet50 for identity verification; they achieved a relatively high recognition accuracy using the CMPD300 dataset [16]. Awad and Hassaballah utilized nasal ridge images in conjunction with the bag-of-words method to address the problem, using accelerated robust features and regions of maximum stability as feature sources, achieving a best accuracy of 93%; however, a drawback was the small number of samples in the dataset [17]. Kaur et al. further applied traditional

features and deep neural networks to nasal pattern analysis, demonstrating that nasal patterns can serve as a reliable biological marker for cattle [18]. However, current literature on cattle muzzles is limited to close-up photographs or manually selected nasal images, with little consideration given to the automatic identification of muzzle locations in the complex environment of open pastures [15,16,17,18].

Subsequently, cattle face detection became a prominent feature in biometric technology [19]. Xu et al. combined RetinaFace-Mobilenet with the ArcFace loss function to design the CattleFaceNet model, which successfully localized cattle faces and performed individual identification, achieving an accuracy of 91.3% [20]; Yang et al. combined RetinaFace with an improved version of FaceNet to thoroughly extract features for cattle face recognition [21]; Qi Yongsheng et al. [22] analyzed cattle face detection in complex backgrounds and adopted a composite dual-branch adaptive attention structure in YOLOv7-tiny, thereby improving the model's detection performance; Gao Jie and Cao Hao [23] incorporated CBAM-style attention, the FReLU function, and CARAFE formal sampling into their analysis, optimizing YOLOv7 accordingly to improve cattle face detection performance.

Some research has been conducted on detection algorithms that combine features from a cow's face and muzzle. Following the approach proposed by Pathak and Prakash, experiments were conducted using YOLOv8 to consider both facial and nasal features together; the proposed method achieved relatively successful recognition on the CMPD300 dataset [24]. However, the aforementioned dataset uses images captured under controlled conditions (with optimal lighting and object visibility), whereas in real-world pasture settings, variations in lighting, cattle posture, and interfering factors such as dust and water marks can hinder the accurate recognition of cattle facial or muzzle features [25].

The issues discussed in this paper constitute the scope of this study. Based on YOLOv8, we have designed the ECL-YOLOv8 model [26] to perform simultaneous open-pasture recognition of cattle faces and muzzles. The contributions of this paper are as follows:

1.The enhanced multi-scale channel attention (E-MSCA) module was incorporated into the backbone network, positioned after the SPPF and at the end of the backbone network. By combining horizontal and vertical strip-depth convolutions and embedding channel attention, it enhances multiscale spatial feature extraction and improves the model's ability to capture features of the cow's face and muzzle [27,28].

2.The C5 layer of the original backbone network and the P5 layer of the neck network are replaced with the C2f-based bottleneck deformable convolution fusion (CBDCN2F) module. By adopting dual-

105 branch feature extraction, the model adapts to deformations caused by cows in different poses, reduces
106 feature loss due to sampling, and simultaneously provides higher-quality input to the E-MSCA module
107 [29,30].

108 3.The Local enhanced channel attention (LECA) module is placed at the outputs of layers P3, P4, and
109 P5 in the Neck network, immediately before the Detect layer. It operates on feature maps at different scales,
110 emphasizing fine-grained texture information in the mouth and muzzle regions to enhance detection
111 capabilities in these areas.

112 **Methods**

113 **Data Collection**

114 This study focuses on the detection of beef cattle faces and muzzles in an open pasture environment; the
115 image data were collected from a beef cattle farm in Xing County, Lüliang City, Shanxi Province. The
116 dataset includes images captured under various conditions, such as different distances, angles, exposures,
117 and obstructions caused by farm facilities. The images were captured using a smartphone, with shooting
118 distances ranging from 0.5 to 4.0 meters when held by hand. Multi-angle images were acquired without
119 disturbing the cattle, covering a variety of scenarios, as shown in Figure 1.



(a) Close up



(b) Obstructions in Livestock Facilities



(c) Long-distance



(d) Eating grass



(e) Exposure



(f) Mud on the Face

Figure 1:Rendering of an open, complex ranch scene.

Dataset Annotation and Segmentation

In this study, the self-built dataset was divided into training, validation, and test sets in a 7:2:1 ratio. The images in the dataset were annotated using LabelImg, an open-source image annotation tool, as shown in Table 2.

Table 2:Results of the dataset split.

Dataset	Number of Images	Number of frames in the "Cow Face" sequence	Number of Muzzle Rims on Cattle
Training Set	2686	4187	3337
Validation Set	767	1249	993
Test Set	385	619	509
Total	3838	6055	4839

Analysis of Target Dimensions in the Face and Muzzle of Cattle

To quantitatively describe the proportions of the cow's face and muzzle, we use the normalized area of the bounding box was calculated as shown in Eq. (1):

$$A_{\text{norm}}=\frac{w_b \text{h}_b}{W_I H_I} \tag{1}$$

In this context,w_b and h_b represent the width and height of the target bounding box, respectively, while W_I and H_I represent the width and height of the original image, respectively.

Table 3：Distribution of Normalized Scale Values for the Facial and Nasal Bounding Boxes of Cattle

Category	Number of frames	Median Normalized Width (Q1–Q3)	Median Normalized Height (Q1–Q3)	Median Normalized Area (Q1–Q3)
Face	6055	0.195 (0.086-0.347)	0.152 (0.073-0.263)	0.0307 (0.0065-0.0827)
Muzzle	4839	0.048 (0.021-0.082)	0.018 (0.010-0.028)	0.00085 (0.00024-0.00213)

134

135 According to the results in Table 3, obtained using statistical methods, the median normalized area of
136 the annotated regions on the cow's face is approximately 36.12 times that of the nasal region, providing a
137 quantitative reference value.

138 **Overall Architecture of ECL-YOLOv8**

139 To address the numerous challenges in cattle face and muzzle recognition in the complex environment of
140 open pastures—such as facial occlusion caused by livestock facilities, muddy faces, and exposure issues—
141 This study proposes ECL-YOLOv8, a model based on an enhanced version of YOLOv8, specifically
142 designed for pasture applications. Unlike approaches that focus solely on local optimization of individual
143 modules, this model systematically reconstructs the original network across three levels: enhancing
144 multiscale features, improving pose adaptability, and strengthening local features. This approach
145 simultaneously improves detection accuracy and inference efficiency while keeping the overall
146 computational complexity within manageable limits.

147 Compared to the baseline model YOLOv8, the key improvements of ECL-YOLOv8 can be
148 summarized as follows:

149 1. Multiscale Spatial and Channel Information Fusion: In the backbone network, E-MSCA is placed
150 after SPPF, at the end of the backbone. E-MSCA combines multiscale spatial feature extraction with
151 channel recalibration methods, which not only suppresses complex background noise buxt also enhances
152 the ability to capture facial and nasal features of cows.

153 2. Enhanced Adaptability to Varied Postures: The C2f convolutions in the backbone's C5 layer and
154 the neck's P5 layer are replaced with CBDCN2F convolutions. The CBDCN2F convolutions in the
155 backbone are primarily used to adapt to facial and nasal deformations caused by cows in different postures,
156 providing higher-quality input to E-MSCA. The CBDCN2F convolution in the neck layer reduces feature

information loss caused by downsampling and improves the model's ability to localize features despite occlusions from livestock facilities, mud on the face, and irregular areas.

3. Enhanced Local Features: An LECA module is inserted before the detection head, acting on feature maps at different scales to further optimize the features fed into the detection head. Through local feature extraction and channel recalibration, fine-grained texture information on the muzzle is highlighted, strengthening the model's ability to detect small-scale features on the muzzle and significantly improving its robustness.

In summary, ECL-YOLOv8 introduces the E-MSCA, CBDCN2F, and LECA modules in the backbone network, neck network, and local enhancement stage, respectively, to construct a hierarchical and functionally complementary feature modeling framework. The entire network architecture is shown in Figure 2.

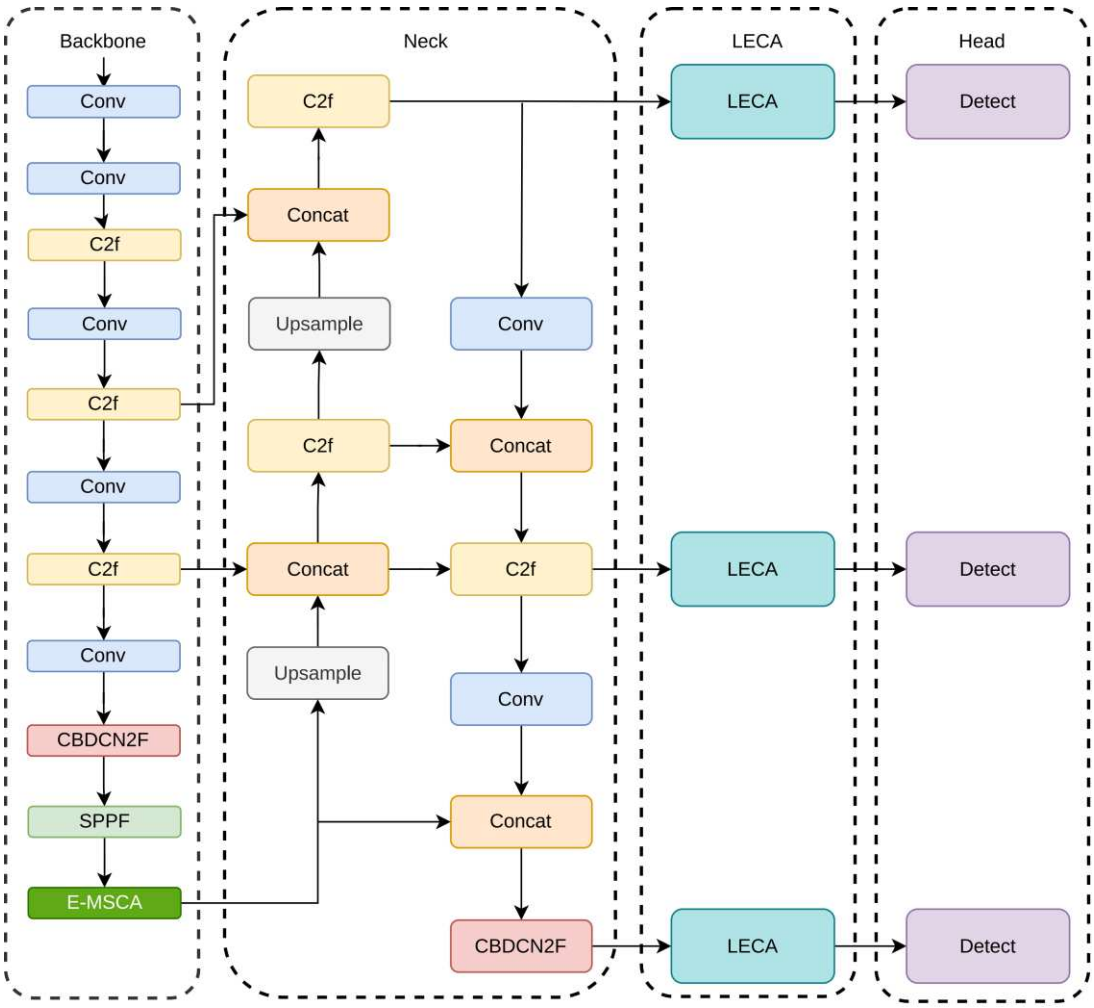


Figure 2:ECL-YOLOv8 Network Architecture Diagram

E-MSCA Module

The E-MSCA module combines multiscale spatial feature extraction with a channel recalibration method to enhance the ability to represent features of the cow's face and muzzle, as shown in Figure 3.

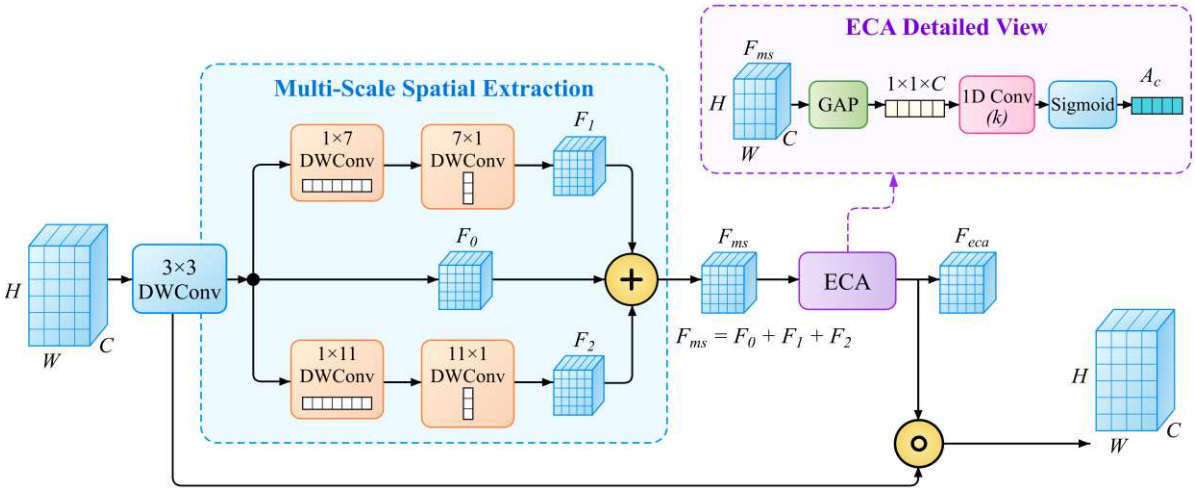


Figure 3:Block Diagram of the E-MSCA Module

First, input the feature $X \in \mathbb{R}^{C \times H \times W}$. After applying a 3×3 deep convolution, local spatial information is extracted to capture the contour of the cow's face and the boundaries of its mouth and muzzle, yielding the base feature F_0, as shown in Eq. (2):

$$F_0 = DWConv_{3\times3}(X) \tag{2}$$

Subsequently, F0 splits into three parallel branches. The middle branch directly passes through the base features; the upper branch uses a 1×7 deep convolution to expand the receptive field horizontally and a 7×1 deep convolution to further aggregate information vertically; the lower branch uses a 1×11 deep convolution to focus on the overall width of the cow's face and the relationships between the muzzle and the surrounding facial areas, as well as an 11×1 deep convolution to model the overall structure of the cow's face from the forehead and eyes to the muzzle, thereby further expanding the spatial receptive field. The outputs of the two convolutional branches are represented by Eq. (3) and Eq. (4), respectively:

$$F_1 = DWConv_{7\times1}(DWConv_{1\times7}(F_0)) \tag{3}$$

$$F_2 = DWConv_{11\times1}(DWConv_{1\times11}(F_0)) \tag{4}$$

By combining horizontal and vertical strip-depth convolutions, the module extracts spatial information across different scales with low parameter overhead. Next, the outputs from the three branches are added element -wise to obtain multiscale fused features,as shown in Eq. (5):

$$F_{\mathrm{ms}}=F_0+F_1+F_2 \tag{5}$$

Next, the F_{ms} is fed into the ECA channel attention module [27]. First, spatial information from each channel is aggregated via global average pooling to generate a channel description vector; then, one-dimensional convolutions are used to model local cross-channel dependencies, and channel weights are generated using the Sigmoid function,as defined in Eq. (6):

$$A_e=\sigma(\mathrm{Conv1d}_k(\mathrm{GAP}(F_{ms}))) \tag{6}$$

Broadcast the channel weights A_c across the spatial dimensions, multiply them element-wise by the multiscale fused features, and recalibrate the feature responses across different channels to obtain,as defined in Eq. (7):

$$F_{\mathrm{eca}}=F_{\mathrm{ms}}\odot A_c \tag{7}$$

Finally, the channel-weighted feature F_{eca} is multiplied element-wise with the base feature F_0, which is passed through the bypass path below, to obtain the module output,as defined in Eq. (8):

$$Y=F_0\odot F_{\mathrm{eca}} \tag{8}$$

Here, $\odot$ denotes element-wise multiplication. The output features retain the same spatial dimensions and number of channels as the base features and are passed to the subsequent network. Through this process, the module enhances the robustness of the cow's face and muzzle by combining local details, multiscale spatial information, and channel importance.

CBDCN2F Module

As shown in Figure 4, the CBDCN2F module employs a structure that incorporates channel expansion, dual-branch feature extraction, and multi-level feature fusion.

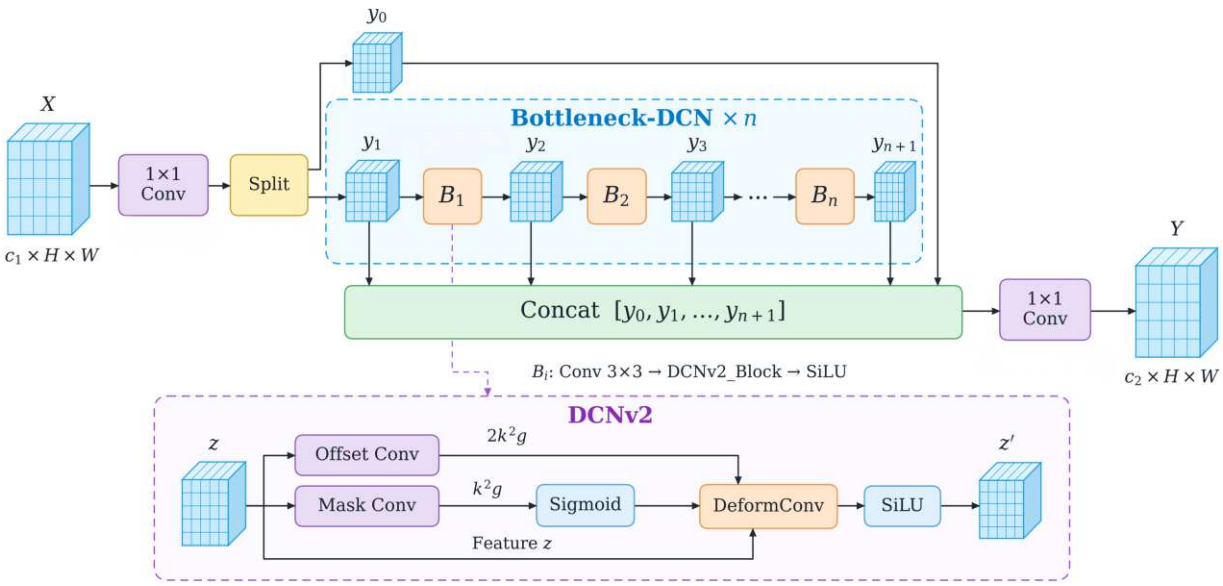


Figure 4:Block Diagram of the CBDCN2F Module

The input feature $X \in \mathbb{R}^{C \times H \times W}$ first undergoes a 1×1 convolution to expand the number of channels to 2c, thereby increasing the feature representation space and enabling the reorganization of information across channels,as defined in Eq. (9):

$$F = \mathrm{Conv}_{1\times1}(X), \quad F \in \mathbb{R}^{2C \times H \times W} \tag{9}$$

Next, F is divided along the channel dimension into fast-branch features F_s and main-branch features F_m, both of which have dimensions of C×H×W [26]. The fast branch directly passes through features, which are used to preserve basic information such as the cow's facial contours, the position of the muzzle, and local textures, while also improving the network's gradient propagation capability.

The main branch consists of n Bottleneck-DCN modules connected in series. By introducing deformable convolutions [29,30] into the Bottleneck modules, the model can adaptively adjust sampling positions based on the actual shapes of the cow's face and the muzzle region, thereby enhancing its ability to adapt to cows in different poses and partial facial occlusions. For small and irregularly shaped muzzle region targets, deformable convolutions provide greater flexibility in aggregating discriminative features such as nostril edges, muzzle contours, and local textures.

Next, the features from the shortcut branch y_0, the initial features from the main branch y_1, and the output features from each level of the Bottleneck-DCN modules $y_2,\ldots,y_{n+1}$ are concatenated along the channel dimension,as defined in Eq. (10) and Eq. (11):

$$F_{cat}=Concat[y_0,y_1,\ldots,y_n,y_{n+1}] \tag{10}$$

231

$$F_{cat}\in\mathbb{R}^{(n+2)c\times H\times W} \tag{11}$$

232

233 This operation preserves both the original structural information and the adaptive features enhanced
234 by deformable convolution. Finally, a 1×1 convolution is used to perform cross-channel fusion on the
235 concatenated features, mapping the number of channels $(n+2)c$ to c_2.The corresponding operations are
236 formulated in Eq. (12) and Eq. (13):

$$Y=\mathrm{Conv}_{1\times1}(F_{cat}) \tag{12}$$

237

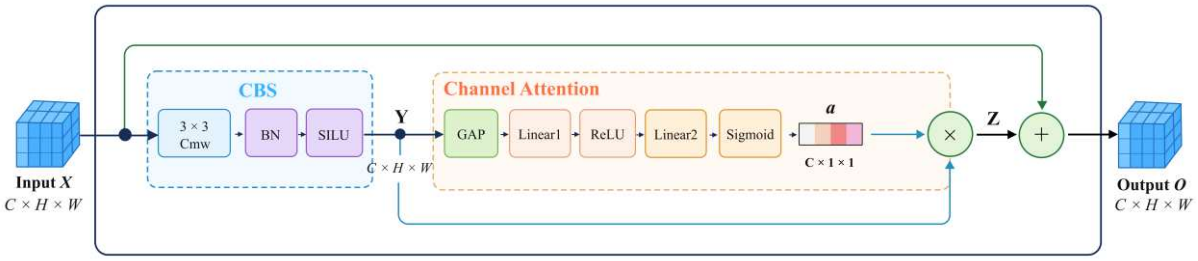
$$Y\in\mathbb{R}^{C\times H\times W} \tag{13}$$

238

239 Therefore, by fusing stable baseline features with features capable of adapting to deformations
240 without altering the spatial resolution of the feature map, CBDCN2F helps improve the detection of cattle
241 faces and small-scale nasal targets in complex pasture environments.

242 **LECA Module**

243 To enhance the discriminative features of cattle faces and muzzles in complex pasture environments, this
244 paper proposes a local enhancement channel attention module, whose structure is shown in Figure 5. The
245 module adaptively enhances the input features through local spatial feature extraction, channel weight
246 learning, and residual fusion.



247

248 Figure 5:LECA Module Structural Diagram

249 The input features are: $X\in\mathbb{R}^{C\times H\times W}$. The module first feeds the input features into the CBS architecture.
250 The CBS consists of a 3×3 convolution, batch normalization, and the SiLU activation function,as defined
251 in Eq. (14):

$$Y=\mathrm{SiLU}(\mathrm{BN}(\mathrm{Conv}_{3\times3}(X))) \tag{14}$$

252

When the stride is 1 and appropriate padding is applied, the spatial dimensions of the output feature map remain unchanged.

In particular, 3×3 convolutions are used to extract local spatial information from the cow's face and muzzle, including the facial contours, the edges of the nostrils, the texture of the muzzle region, and hair boundaries. The BN reduces variations in feature distributions across different batches, thereby improving training stability; the SiLU activation function introduces nonlinear expressive power, enabling the network to more fully capture local features in complex backgrounds.

Next, the local features Y are fed into the channel attention branch. Global average pooling is used to aggregate spatial information for each channel, yielding a channel description vector,as defined in Eq. (15):

$$s_c=\frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W} X_c(i,j) \tag{15}$$

In this case, $s\in\mathbb{R}^{C\times 1\times 1}$. This operation compresses the spatial features of each channel into a single global response value, which is used to describe the overall semantic information contained in that channel.

Based on this, a multilayer perceptron consisting of two fully connected layers is used to learn the dependencies between channels. The calculation process is formulated in Eq. (16) and Eq. (17):

$$a=\mathrm{ReLU}(W_1 s+b_1) \tag{16}$$

$$\alpha=\sigma(W_2 a+b_2) \tag{17}$$

Here, W_1 and W_2 represent two linear transformations, respectively; ReLU is used to enhance nonlinear representational capacity, and the sigmoid function is used to constrain the channel weights to the range of 0 to 1, ultimately yielding the channel attention weights,as defined in Eq. (18):

$$\alpha\in\mathbb{R}^{C\times 1\times 1} \tag{18}$$

For cow face and muzzle detection, this process automatically determines the importance of different channels based on the current input features, assigning higher weights to channels containing relevant information—such as the cow's facial outline, the edge of the muzzle, and the texture of the nostrils—while minimizing the impact of background, shadows, and irrelevant hair details.

Next, the channel weight α is broadcast across the spatial dimensions and multiplied element-wise with the local features Y to obtain the channel-weighted features,as defined in Eq. (19):

$$Z=Y\odot\alpha \tag{19}$$

Here, $\odot$ denotes element-wise multiplication. This operation does not alter the spatial dimensions of the feature map but redistributes the response intensity across each channel, thereby highlighting features relevant to the detection of a cow's face and muzzle.

Finally, add the residual to the weighted features Z and the input features X,as defined in Eq. (20):

$$O=X+Z,\ O\in\mathbb{R}^{C\times H\times W} \tag{20}$$

Residual connections preserve the original details in the input features, prevent channel attention from excessively suppressing weak target features, and improve gradient propagation in the network. For the muzzle—a small-area target with limited textural information—the residual structure helps preserve fine-grained details such as the nostrils and the boundaries of the muzzle.

In summary, the LECA module extracts local spatial features using CBS, establishes channel dependencies via GAP and MLP, adaptively recalibrates the features using channel weights, and preserves the original information through residual connections. Before feeding the data into the detection head, it enhances the expressive power of effective features in the cow's face and muzzle, with a particular emphasis on fine-grained textural information in the muzzle.

Results

Experimental Setup

All models were trained and tested using the same hardware, dataset split, and evaluation metrics. The main configurations are shown in Table 4.

Table 4:Experimental Environment and Training Parameters

Category	Project	Settings
Experimental Environment	Operating System	Ubuntu 22.04.5 LTS
	GPU	NVIDIA GeForce RTX 4090 (24 GB)
	Programming Languages	Python 3.12.11
	Deep Learning Frameworks	PyTorch 2.9.1
	CUDA Version	CUDA 13.0
Training Parameters	Optimizer	AdamW
	Initial learning rate	0.001
	Batch Size	16
	Input Image Dimensions	640 × 640
	Number of Training Rounds	300

Evaluation Criteria

To conduct a comprehensive performance analysis of the model being used, the following metrics should be included: precision (P), recall (R), mean average precision (mAP), number of parameters, computational load (GFLOPs), and inference speed (FPS).

Of the two metrics listed, Precision refers to the proportion of correctly identified targets, while Recall is an estimate of the proportion of actual targets that are detected. The corresponding formulas are defined in Eq. (21)and Eq. (22):

$$P=\frac{TP}{TP+FP} \tag{21}$$

$$R=\frac{TP}{TP+FN} \tag{22}$$

The so-called average accuracy is calculated based on the Precision-Recall curves corresponding to different confidence thresholds,as defined in Eq. (23):

$$AP=\int_0^1 P(R)\,dR \tag{23}$$

Metrics related to multi-class classification should be expressed as the mean precision,which is defined in Eq. (24):

$$mAP=\frac{1}{N}\sum_{i=1}^{N} AP_i \tag{24}$$

Here,N represents the number of detection classes. The paper explains two mAP metrics: mAP50, calculated using an IoU of 0.5, and mAP50:95, calculated using IoU values ranging from 0.5 to 0.95 (in 0.05 increments).

In addition, the number of model parameters, the volume of floating-point operations, and the number of frames detected per second are used as metrics to assess the model's computational complexity and real-time performance.

Analysis of Muzzle Detection Performance at Different Target Scales

To analyze the model’s ability to detect cow muzzle targets at different scales, the test set targets were divided into three groups—Small, Medium, and Large—based on the proportion

of the actual bounding box area to the image area, and the recall rates were calculated for each group. The results are shown in Table 5.

Table 5:Object Recall Rates for Cattle Muzzles, Grouped by Normalized Area Quantiles

Model	Small Group Recall (%)	Medium Group Recall (%)	Large Group Recall (%)
YOLOv8n	26.90 (46/171)	95.24 (160/168)	87.65 (149/170)
ECL-YOLOv8	52.63 (90/171)	97.02 (163/168)	85.29 (145/170)
Improvement	+25.73	+1.79	–2.35

As shown in the results in Table 5, ECL-YOLOv8's recall rate for the Small group increased from 26.90% to 52.63%, representing an increase of 25.73 percentage points; for the Medium group, it rose by 1.79 percentage points; and for the Large group, it decreased by 2.35 percentage points.

Ablation Experiment

Table 6:ECL-YOLOv8 Ablation Experiment Results

Config	P	R	mAP50	mAP50:95	Face AP50	Muzzle AP50	Face AP50:95	Muzzle AP50:95	Params (M)	GFLOPs	FPS
YOLOv8	91.6	77.5	84.3	54.2	91.6	77.1	63.9	44.5	3.01	8.1	211.11
+A	93.2	79.5	86.1	54.9	90.6	81.6	62.7	47.1	2.70	6.80	192.45
+B	90.1	81.8	86.9	55.1	91.7	82.0	63.2	47.1	2.83	10.7	161.42
+C	90.6	80.7	86.0	55.3	91.1	80.8	63.1	47.4	3.60	12.8	153.82
+A+B	91.9	80.5	87.0	55.9	92.1	81.8	64.6	47.2	3.16	11.2	145.98
+A+B+C (ECL)	92.6	81.7	87.4	56.3	92.5	82.4	64.7	48.0	3.91	13.0	131.52

To systematically evaluate and quantify the contribution of each module in this complex model to its overall performance, we conducted ablation experiments on the YOLOv8 model; the results are shown in Table 6. The baseline YOLOv8 model achieved a precision of 91.6%, a recall of 77.5%, an mAP50 of 84.3%, and an mAP50-95 of 54.2%. This study introduces three new modules for cow face and muzzle

detection: the E-MSCA module, the CBDCN2F module, and the LECA module, all of which demonstrate significant advantages. The E-MSCA module significantly enhances the feature representation capability for cow faces and muzzles through multiscale spatial feature extraction. Although the introduction of the E-MSCA module caused a slight decrease in the mAP50 and mAP50-95 for bovine faces, the recall rate (79.5%), as well as the mAP50 (81.6%) and mAP50-95 (82.7%) for bovine muzzles, all increased significantly, thereby enhancing the model's detection capabilities for bovine faces and muzzles. The CBDCN2F module, through systematic designs such as dual-branch feature extraction, channel expansion, and channel fusion, enables the model to adaptively adjust sampling positions, enhancing its adaptability to complex conditions. This significantly improves the accuracy and robustness of cow face and muzzle detection, although precision decreases slightly; recall increases to 81.8%, with overall mAP50 and mAP50-95 reaching 86.9% and 55.1%, respectively. The LECA module further optimizes the model before feeding it into the detection head, thereby enhancing overall performance. While precision decreased slightly, recall increased to 80.7%, mAP50 rose to 86.0%, and mAP50-95 increased to 55.3%. Upon introducing E-MSCA and CBDCN2F, the two modules complemented each other. After the CBDCN2 module replaced the C5 layer of the backbone and the P5 layer of the neck, it not only provided higher-quality input to E-MSCA but also reduced the loss of feature information caused by sampling. The results showed that the overall mAP50 and mAP50-95 reached 87.0% and 55.9%, respectively. When the three modules were used in combination, the model achieved optimal performance, with mAP50 and mAP50-95 reaching 87.4% and 56.3%, respectively—a 3.1% increase in mAP50 compared to the unmodified model. These results indicate that the combined use of these modules can significantly improve the accuracy of cow face and muzzle detection.

A Comparison of the Performance of Different Object Detection Algorithms

To conduct a comparative analysis of the proposed model and various commonly used convolutional neural networks in terms of object detection performance, the proposed algorithm, ECL-YOLOv8, was tested against YOLOv5, YOLOv8, YOLOv9, YOLOv11, and YOLOv12 under identical conditions. The experimental results are shown in Table 7.

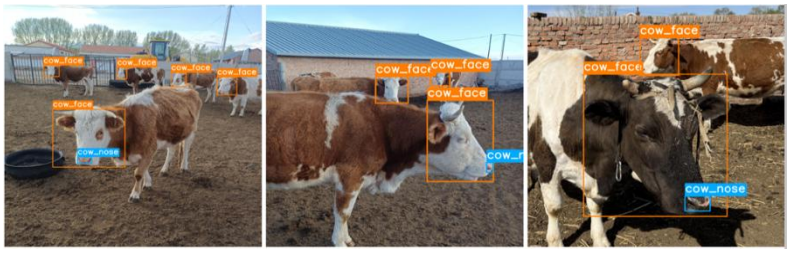
Table 7:Performance Comparison of Different Models in Recognition Tasks

Model	Cattle Face mAP50 (%)	Cattle Muzzle mAP50 (%)	mAP50 (%)	mAP50:95 (%)	Precision (%)	Recall (%)	Number of participants (M)	GFLOPs
YOLOv5	91.1	77.6	84.3	52.8	93.0	78.3	2.50	7.1
YOLOv8	91.6	77.1	84.3	54.2	91.6	77.5	3.01	8.1
YOLOv9	91.5	78.5	85.0	53.4	93.1	79.0	1.97	7.6
YOLO11	91.1	76.6	83.9	52.4	91.5	78.8	2.58	6.3
YOLO12	90.9	78.3	84.6	54.3	93.3	76.0	2.54	6.3
ECL-YOLOv8	**92.5**	**82.4**	**87.4**	**56.3**	**92.6**	**81.7**	3.91	13.0

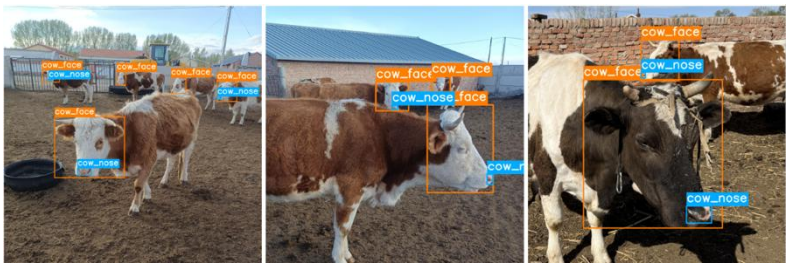
As shown in Table 7, the ECL-YOLOv8 model described in this paper demonstrates good overall performance when simultaneously detecting cattle faces and muzzles. For cattle faces, the mAP50, muzzle mAP50, mAP50, mAP50:95, and recall of 92.5%, 82.4%, 87.4%, 56.3%, and 81.7%, respectively—all of which are the highest values in the current model comparison. Compared to the baseline model YOLOv8, ECL-YOLOv8 shows increases of 3.1 and 2.1 percentage points in mAP50 and mAP50:95, respectively, while precision and recall also increase by 1.0 and 4.2 percentage points, respectively. Here, the mAP50 for the cow's face improved by 0.9 percentage points, while the mAP50 for the muzzle improved by 5.3 percentage points, indicating that the model's optimization for the cow's face did not compromise the recognition performance for the cow's muzzle. From an experimental perspective, although the new method incurs some computational overhead, it significantly improves the quality of joint detection of cattle faces and muzzles. In particular, it enables more accurate localization of the cattle muzzle, laying the foundation for individual identification.

Visual Analysis of Test Results

To fully demonstrate the recognition capabilities of the improved model, we will select several examples of cow faces for experimental analysis. The data set itself is highly complex and difficult to process, making it well-suited for evaluating the model's performance. The results for both the improved ECL-YOLOv8 and the original YOLOv8 are shown in Figure 6.



YOLOv8



ECL-YOLOv8

Figure 6:Comparison of Detection Results Between YOLOv8 and ECL-YOLOv8

Discussion

This study investigates the problem of accurately recognizing cattle faces and muzzle regions in the complex environment of open pastures and proposes an improved detection model based on the YOLOv8 architecture—ECL-YOLOv8. This model incorporates E-MSCA after the SPPF module, replaces the C2f in the C5 layer of the main branch and the P5 layer of the neck branch with CBDCN2F, and inserts LECA before the P3, P4, and P5 detection heads, thereby constructing a clear and complementary feature modeling system. Specifically, the multiscale design of E-MSCA enables better detection of targets of varying sizes; CBDCN2F deformable convolutions enhance the model's ability to extract features from cattle targets in various poses; LECA key-channel optimization enables clearer representation of small muzzle regions, aiding in the detection of cattle faces and muzzles. These improvements collectively strengthen the model's ability to represent cattle faces and muzzles across different scales and discriminative dimensions. The core contribution of this study lies in the fact that, compared to traditional object detection models such as YOLOv9, YOLOv11, and YOLOv12, ECL-YOLOv8 demonstrates significant advantages, particularly in snout detection capabilities and adaptability in complex backgrounds. These advantages are primarily attributable to three targeted structural optimizations: a multi-scale design that enables better perception of objects, feature extraction capabilities based on variable convolutions, and localized enhancement of the muzzle region. These optimizations enhance the model's detection capabilities and its ability to capture local details.

Compared to existing improved YOLO methods, ECL-YOLOv8 achieves better performance on key metrics, including precision, recall, mAP50, and mAP50-95, and demonstrates greater robustness under challenging conditions such as complex environments, occlusions caused by livestock facilities, and exposure variations. Specifically, the E-MSCA module enhances multiscale feature extraction capabilities; the CBDCN2F convolutional layer improves feature extraction for cattle in various poses; and LECA

enhances the detection of the muzzle region in front of the head without affecting the cattle's facial recognition. Furthermore, systematic ablation experiments and comparative analyses indicate that ECL-YOLOv8 not only maintains high detection accuracy but also balances computational efficiency. Its robustness has been validated on datasets featuring complex environments (such as varying exposure levels or muddy faces), fully demonstrating the model's immense potential for practical applications.

Although the proposed model for detecting cattle faces and muzzles regions offers significant advantages in terms of recognition performance, some limitations remain. For example, while the current model's detection frame rate (FPS) meets real-time processing requirements, its inference speed is still slower than that of baseline models. Further optimization is still needed for high-speed motion or resource-constrained application scenarios. Additionally, the model's large number of parameters poses challenges for deployment on devices with limited computational capabilities. Future research could focus on technical areas such as model compression, lightweight design, and inference acceleration to further enhance the model's efficiency and practicality. This would better support real-time monitoring and intelligent operations on actual ranches, providing more reliable technical support for smart agriculture applications.

Conclusion

This study found that, in the complex environment of open pastures, the ECL-YOLOv8 model effectively improved the detection performance of cattle faces and muzzles, with particularly notable improvements in muzzle detection. Experimental results on our self-built dataset show that our model achieved a 4.2 percentage point increase in recall compared to the baseline YOLOv8; specifically, the average precision for muzzle detection reached 82.4%, representing a 5.3 percentage point improvement over the baseline. A comprehensive analysis indicates that the ECL-YOLOv8 model effectively improves the detection performance of cattle faces and snouts in the study setting, laying the foundation for individual cattle identification in complex pasture environments. Future work should further validate the model's generalization capabilities across different pastures or independent datasets.

List of abbreviations

E-MSCA Enhanced multi-scale channel attention

CBDCN2F C2f-based bottleneck deformable convolution fusion

LECA Local enhanced channel attention

Figure Legends

Figure 1:Rendering of an open, complex ranch scene.

Figure 2:ECL-YOLOv8 Network Architecture Diagram

Figure 3:Block Diagram of the E-MSCA Module

Figure 4:Block Diagram of the CBDCN2F Module

Figure 5:LECA Module Structural Diagram

Figure 6:Comparison of Detection Results Between YOLOv8 and ECL-YOLOv8

Additional Files Legends

Not applicable.

Declarations

- **Ethics approval and consent to participate**

  Not applicable

- **Consent for publication**

  Not applicable

- **Availability of data and materials**

The datasets generated and analyzed during the current study are not publicly available due to data ownership restrictions but are available from the corresponding author on reasonable request.

- **Competing interests**

The authors declare that they have no competing interests.

- **Funding**

This research was supported by the Science and Technology Innovation Enhancement Project (Grant No. CXGC2026058), the Basic Research Program of Shanxi Province (Youth Program) (Grant No. 202303021222066), the Young and Middle-aged Elite Innovative Talent Training Program (Grant No. SXAUKY2024002), the Key Research and Development Program of Lvliang City (Grant Nos. 2025NY17 and 2025NY18), and the Graduate Quality Engineering Project of Shanxi Agricultural University (Grant Nos. 2025YZLGC033 and 2025JG056).

The funding bodies had no role in the design of the study, collection, analysis, and interpretation of data, or in writing the manuscript.

Authors' contributions

JL:Conceptualization,Methodology,Software,Data curation,Formal analysis,Visualization,Writing--original draft.

SZ:Investigation,Data curation,Validation,Writing--review and editing.

YL:Investigation,Data curation,Validation.

HZ:Software,Formal analysis,Visualization.

ZJ:Investigation,Resources.

YD:Data curation,Validation.

RL:Writing--review and editing.

JH:Conceptualization,Supervision,Project administration,Funding acquisition,Writing--review and editing.

All authors contributed to the article and approved the submitted version.

- **Acknowledgements**

The authors thank all members of the research group for their assistance in data collection, annotation, and experimental support.

References

1. FAO. The future of food and agriculture: Trends and challenges. Rome: Food and Agriculture Organization of the United Nations; 2017.

2. Kleen JL, Guatteo R. Precision livestock farming: what does it contain and what are the perspectives? Animals. 2023;13(5):779. doi:10.3390/ani13050779.

3. Neethirajan S. The role of sensors, big data and machine learning in modern animal farming. Sensing and Bio-Sensing Research. 2020;29:100367. doi:10.1016/j.sbsr.2020.100367.

4. Meng H, Zhang L, Yang F, et al. Livestock biometrics identification using computer vision approaches: a review. Agriculture. 2025;15(1):102. doi:10.3390/agriculture15010102.

5. Barry B, Gonzales-Barron U, McDonnell K, Butler F, Ward S. Using muzzle pattern recognition as a biometric approach for cattle identification. Computers and Electronics in Agriculture. 2007;57(1):57-65. doi:10.1016/j.compag.2007.01.012.

6. Wang Z, Zhang Y, Wang X, et al. Cattle face recognition based on deep learning. Computers and Electronics in Agriculture. 2021;187:106292.

7. Li D, Zhang Y, Wang X, et al. Cattle identification based on multiple feature decision fusion of cattle faces, muzzle patterns and ear tags. Scientific Reports. 2024;14:24704. doi:10.1038/s41598-024-76718-x.

8. Kamilaris A, Prenafeta-Boldú FX. Deep learning in agriculture: a survey. Computers and Electronics in Agriculture. 2018;147:70-90. doi:10.1016/j.compag.2018.02.016.

9. Zou Z, Shi Z, Guo Y, et al. Object detection in 20 years: a survey. Proceedings of the IEEE. 2023;111(3):257-276. doi:10.1109/JPROC.2023.3238523.

10. Ren S, He K, Girshick R, Sun J. Faster R-CNN: towards real-time object detection with region proposal networks. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2017;39(6):1137-1149. doi:10.1109/TPAMI.2016.2577031.

11. Carion N, Massa F, Synnaeve G, et al. End-to-end object detection with transformers. In: Proceedings of the 16th European Conference on Computer Vision (ECCV 2020). Cham: Springer; 2020. p. 213-229. doi:10.1007/978-3-030-58452-8_13.

12. Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: unified, real-time object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE; 2016. p. 779-788. doi:10.1109/CVPR.2016.91.

13. Terven J, Cordova-Esparza DM. A comprehensive review of YOLO architectures in computer vision: from YOLOv1 to YOLOv8 and YOLO-NAS. Machine Learning and Knowledge Extraction. 2023;5(4):1680-1716. doi:10.3390/make5040080.

14. Wang CY, Bochkovskiy A, Liao HYM. YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023. p. 7464-7475. doi:10.1109/CVPR52729.2023.00724.

15. Kumar S, Singh SK. Cattle recognition: a new frontier in visual animal biometrics research. Proceedings of the National Academy of Sciences, India Section A: Physical Sciences. 2020;90(4):689-708. doi:10.1007/s40010-019-00610-x.

16. Hojaeipour M, Sadeghi M, Hajiaghajani M, et al. Cattle identification using muzzle images based on deep learning. Computers and Electronics in Agriculture. 2021;188:106344. doi:10.1016/j.compag.2021.106344.

17. Awad AI, Hassaballah M. Image-based identification of individual cattle via muzzle print pattern recognition. Multimedia Tools and Applications. 2019;78:30645-30669. doi:10.1007/s11042-019-07859-1.

18. Kaur A, Kumar M, Jindal MK. Cattle identification with muzzle pattern using computer vision technology: a critical review and prospective. Soft Computing. 2022;26:4771-4790. doi:10.1007/s00500-022-06935-x.

19. Awad AI. From classical methods to animal biometrics: a review on cattle identification and tracking. Computers and Electronics in Agriculture. 2016;123:423-435. doi:10.1016/j.compag.2016.03.014.

20. Xu B, Wang W, Guo L, et al. CattleFaceNet: a cattle face identification approach based on RetinaFace and ArcFace loss. Computers and Electronics in Agriculture. 2022;193:106675. doi:10.1016/j.compag.2021.106675.

21. Yang H, Zhang Y, Zhang Y, et al. Cattle face recognition based on improved FaceNet. Computers and Electronics in Agriculture. 2022;193:106688. doi:10.1016/j.compag.2022.106688.

22. Qi YS, Jiao J, Bao TF, et al. Cattle face detection algorithm in complex scenes using adaptive attention mechanism. Transactions of the Chinese Society of Agricultural Engineering. 2023;39(14):173-183.

23. Gao J, Cao H. Research on cow face recognition method based on improved YOLO v7. Journal of Xinyang Agriculture and Forestry University. 2024;34(3):125-130.

24. Pathak PD, Prakash S. Attention-based multi-modal robust cattle identification technique using deep learning. Computers and Electronics in Agriculture. 2025;238:110747. doi:10.1016/j.compag.2025.110747.

25. Noe SM, Zin TT, Kobayashi I, et al. Optimizing black cattle tracking in complex open ranch environments using YOLOv8 embedded multi-camera system. Scientific Reports. 2025;15:6820.

26. Jocher G, Chaurasia A, Qiu J. Ultralytics YOLOv8 (Version 8.0.0) [software]. 2023. Available from: https://github.com/ultralytics/ultralytics.

27. Wang Q, Wu B, Zhu P, et al. ECA-Net: efficient channel attention for deep convolutional neural networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE; 2020. p. 11534-11542. doi:10.1109/CVPR42600.2020.01155.

28. Lin TY, Dollár P, Girshick R, et al. Feature pyramid networks for object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017. p. 2117-2125. doi:10.1109/CVPR.2017.106.

29. Dai J, Qi H, Xiong Y, et al. Deformable convolutional networks. In: Proceedings of the IEEE International Conference on Computer Vision. 2017. p. 764-773. doi:10.1109/ICCV.2017.89.

30. Zhu X, Hu H, Lin S, Dai J. Deformable ConvNets v2: more deformable, better results. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019. p. 9308-9316. doi:10.1109/CVPR.2019.00954.