Journal of Electronic Research and Application, 2026, Volume 10, Issue 8
http://ojs.bbwpublisher.com/index.php/JERA
Online ISSN: 2208-3510
Print ISSN: 2208-3502

Research on Plateau Cattle and Sheep Detection Based on QHCSM-GAN Image Completion and Improved YOLO-OC Network

Xin Zhang

Qinghai Minzu University, Xining, Qinghai, China

Abstract: Highland animal husbandry is a core pillar of Qinghai Province's agricultural and pastoral economy, and computer vision-based intelligent monitoring is a key approach to achieving refined and digital management of highland pastures. However, the complex environment at high altitudes presents challenges such as variable lighting, complex terrain, frequent target occlusion, and blurred imaging, leading to severe degradation in the quality of cattle and sheep images collected in the field and the loss of effective features. Furthermore, existing publicly available livestock datasets and general detection models have poor adaptability to open grazing scenarios on high plateaus, exhibiting shortcomings such as low detection accuracy and weak generalization ability. To address these issues, this paper proposes an integrated intelligent detection framework that combines image completion and target detection optimization. First, a dedicated Qinghai cattle and sheep dataset (QH-Livestock) covering various complex plateau scenes is constructed. Second, a generative adversarial network (QHCSM-GAN) for missing images of plateau cattle and sheep is designed, relying on a multi-path feature fusion encoder and a theoretically reliable discriminator to achieve degraded image restoration and missing feature completion. Finally, using YOLOv8n as the baseline model, a lightweight and high-precision YOLO-OC detection model is constructed by introducing full-dimensional dynamic convolution (ODConv) and convolutional block attention module (CBAM), enhancing multi-scale feature extraction capabilities and anti-interference capabilities against complex backgrounds. Experimental results show that QHCSM-GAN outperforms mainstream image completion algorithms in terms of structural similarity (SSIM), peak signal-to-noise ratio (PSNR), and learned perceptual patch similarity (LPIPS). The proposed YOLO-OC model achieves an accuracy of 84.6%, a recall of 81.2%, and an mAP50 of 82.5% on the QH-Livestock test set, with an inference speed of 109 FPS. This method effectively solves the core problems of poor image quality and insufficient detection accuracy of cattle and sheep targets in complex high-altitude scenes, and can provide reliable technical support for intelligent monitoring of high-altitude pastures and digital transformation of animal husbandry.

Keywords: Plateau animal husbandry; Image completion; Deep learning; Object detection; YOLO-OC; Intelligent monitoring

Online publication: Sept 4, 2026

1.Introduction

Traditional manual inspection methods suffer from drawbacks such as low efficiency, high labor costs, and the inability to conduct large-scale, all-weather monitoring, making them unsuitable for the needs of modern ranch management. Intelligent sensing technologies based on deep learning and computer vision enable non-contact, real-time detection and status monitoring of cattle and sheep, and have become a core technological direction for the intelligent upgrading of plateau animal husbandry.

Current mainstream object detection algorithms have achieved good results in standardized, enclosed livestock farms with simple scenes and controllable lighting. However, their applicability is severely limited in open grazing scenarios on plateaus. First, the extreme environment of high altitude, low temperature, frequent sandstorms, and strong sunlight on plateaus leads to instability in image acquisition equipment, resulting in widespread degradation issues such as overexposure, underexposure, blurring, and partial occlusion in acquired images, causing a significant loss of key features of cattle and sheep. Second, the background of plateau grasslands, mountains, and shadows blends closely with the color of cattle and sheep, and the large scale, dispersed distribution, and varied postures of grazing cattle and sheep significantly increase the difficulty of accurate detection. Finally, existing publicly available livestock datasets are mostly built based on standardized inland farming scenarios, lacking labeled samples from extreme plateau environments. This makes general detection models prone to false positives and false negatives in plateau pasture scenarios, resulting in extremely poor generalization performance.

To address the aforementioned industry pain points, existing research primarily focuses on improvements in two dimensions: data augmentation and network structure optimization. At the data optimization level, traditional image inpainting methods can only handle minor image imperfections and cannot repair feature loss caused by large-area occlusion or strong light degradation. While traditional Generative Adversarial Networks (GANs) possess image generation capabilities, they suffer from problems such as pattern collapse, gradient instability, and unquantifiable uncertainty in generated results, easily leading to texture distortion and false targets [1]. At the detection model level, mainstream YOLO series algorithms, with their fixed convolutional kernels and single feature enhancement mechanisms, cannot dynamically adapt to multi-scale target changes in high-altitude environments, struggle to effectively suppress complex background interference, and their detection accuracy falls short of actual monitoring needs [2]. Therefore, there is an urgent need to construct an integrated solution for high-quality data construction, adaptive image inpainting, and scene-based detection optimization adapted to high-altitude scenarios.

The main research contributions of this paper are as follows: (1) A high-precision labeled QH-Livestock cattle and sheep dataset containing 77,506 images was constructed, covering typical difficult scenarios such as strong light, weak light, occlusion, and distant small targets, filling the gap in the high-altitude livestock-specific dataset; (2) The QHCSM-GAN image completion algorithm was proposed, which quantifies the uncertainty of image generation through a multi-path feature fusion coding structure and a reliable discriminator based on evidence theory, and realizes high-fidelity restoration of high-altitude degraded cattle and sheep images; (3) Based on the YOLOv8n baseline model, a lightweight YOLO-OC detection model was designed by integrating ODConv full-dimensional dynamic convolution and CBAM attention mechanism, which takes into account multi-scale feature extraction capability, background anti-interference capability and real-time inference performance [3,4]; (4) Through multiple sets of comparative experiments, ablation experiments and cross-domain generalization experiments, the effectiveness, robustness and universality of

the algorithm in this paper were fully verified.

The structure of this paper is as follows: Chapter 2 reviews the current research status in the fields of image completion and livestock target detection, and analyzes the shortcomings and research gaps of existing technologies; Chapter 3 details the construction process of the QH-Livestock dataset and the principle of the QHCSM-GAN image completion algorithm; Chapter 4 elaborates on the optimization strategy of the YOLO-OC model structure, experimental settings and result analysis; Chapter 5 summarizes the research results of the entire paper, analyzes the limitations of the research and looks forward to future research directions.

2.Related work

2.1. Research on general target detection based on YOLO series

Single-stage YOLO algorithms, with their advantages of end-to-end inference, strong real-time performance, and excellent balance between accuracy and speed, have become the mainstream algorithms for real-time object detection in industrial scenarios. From the first generation YOLOv1 to the latest YOLOv12, the model has been continuously iterated and optimized in terms of feature fusion mechanism, detection head structure, and training strategy. Early YOLO models had problems such as weak small object detection capability and unbalanced multi-scale feature fusion. Subsequent versions have effectively improved multi-scale object detection performance by introducing modules such as Cross-Stage Local (CSP) structure, Path Aggregation Network (PANet), and decoupled detection head [5,6]. Among them, YOLOv8n, as a lightweight representative model, uses the C2f module to replace the traditional C3 module, optimizes gradient propagation and feature fusion effects, and combines anchorless prediction and classification-regression decoupled branches, achieving an excellent balance between lightweight deployment and detection accuracy, and is widely used in various mobile detection scenarios.

While YOLO models perform well in general scenarios, their fixed convolutional kernels and single feature enhancement mechanisms make them difficult to adapt to the extremely complex high-altitude environments. General-purpose YOLO models cannot dynamically adapt to the large-scale variations in cattle and sheep targets on the plateau, lack targeted suppression capabilities for complex background noise, and are prone to false detections such as missed detections of small distant targets and target-background confusion, severely hindering their practical application in intelligent monitoring scenarios on plateau pastures.

2.2. Current status of research on livestock target detection

With the rapid development of smart agriculture technology, deep learning object detection algorithms have been widely applied to tasks such as livestock and poultry monitoring, counting, and behavior recognition. Existing livestock detection research mostly focuses on standardized farm scenarios with controlled environments, targeting pigs, poultry, and penned cattle and sheep. Researchers have optimized models such as Faster R-CNN, Mask R-CNN, YOLOv5, and YOLOv8 to address issues like dense occlusion and small object detection. This type of research relies on experimental scenarios with simple backgrounds, stable lighting, and uniform target scale, resulting in highly adaptable algorithms, but they cannot be directly transferred to open grazing scenarios.

In the field of image data optimization, traditional image inpainting methods based on partial differential equations, total variational models, and block matching can only repair simple image defects such as minor

scratches and slight blurring, and cannot handle severe image degradation problems such as large-area occlusion and overexposure in high-altitude scenes. Deep learning inpainting methods based on GANs have become the mainstream solution for agricultural image restoration due to their powerful feature modeling and data generation capabilities. However, traditional GAN models suffer from problems such as unstable training gradients, mode collapse, and uncontrollable generation results, leading to texture distortion, structural breaks, and background artifacts in restored cattle and sheep images. While existing attention-enhanced GAN models can strengthen the ability to focus on target features, they lack uncertainty quantification mechanisms and adaptive feature fusion strategies specifically designed for high-altitude scenes, resulting in significant shortcomings in restoration effectiveness.

In summary, the current field of intelligent livestock monitoring lacks dedicated datasets and targeted algorithm optimization schemes adapted to open grazing scenarios on plateaus, and has not yet formed a complete technical system integrating "data augmentation-image restoration-precise detection". This is the core research gap that this paper focuses on addressing.

3.Research methods

3.1. Construction of the QH-livestock plateau cattle and sheep dataset

To address the issues of scarce, limited, and poor-quality data in plateau livestock farming scenarios, this paper constructs a dedicated QH-Livestock plateau cattle and sheep dataset, using a typical plateau pastoral area in Xia Dawu Township, Maqin County, Golog Tibetan Autonomous Prefecture, Qinghai Province as the data collection region. The data collection area covers diverse scenarios including open natural pastures, intensive breeding areas, and herders' courtyards. High-definition monitoring equipment at multiple locations was used to collect images of cattle and sheep activities under different seasons, time periods, and weather conditions. Additionally, a small number of edge scene samples from publicly available datasets were supplemented to enrich the diversity of target poses, scales, and scenarios (See **Figure 1**).

Figure 1. Image data collected from the pastures in Qinghai.

The dataset contains 77,506 high-precision labeled RGB images, categorized into cattle and sheep targets. To unify data distribution and improve sample quality, the original images underwent standardized preprocessing: uniformly cropped and scaled to 640 × 640 pixels; adaptive histogram equalization was used for illumination correction to eliminate overexposure in strong light and underexposure in weak light; Gaussian filtering was used to suppress wind noise and equipment salt-and-pepper noise; and severely

blurred, distorted, or targetless invalid samples were removed. The dataset is randomly divided into training, validation, and test sets in an 8:1:1 ratio to ensure balanced sample distribution across various scenes, scales, and lighting conditions. Three specialized test subsets, small targets, strong light, and occlusion, are also included to accurately validate the model's detection performance in challenging high-altitude scenarios. Due to privacy concerns regarding herders, this dataset is not currently open-source.

3.2. Design of QHCSM-GAN image completion algorithm

To address the issues of strong light degradation, target occlusion, and feature loss in images of cattle and sheep on the plateau, this paper proposes the QHCSM-GAN image completion algorithm. The overall structure consists of a multi-path feature fusion encoder, a decoding module, and an evidence-theoretic trusted discriminator. The algorithm restores details in degraded images through multi-scale feature fusion and relies on evidence theory to quantify generation uncertainty, effectively suppressing artifacts and distortion. **Figure 2** shows the overall network architecture of QHCSM-GAN, fully illustrating the multi-path encoding branches, the CGA-fusion attention fusion module, the decoding output, and the trusted discriminator structure [7].

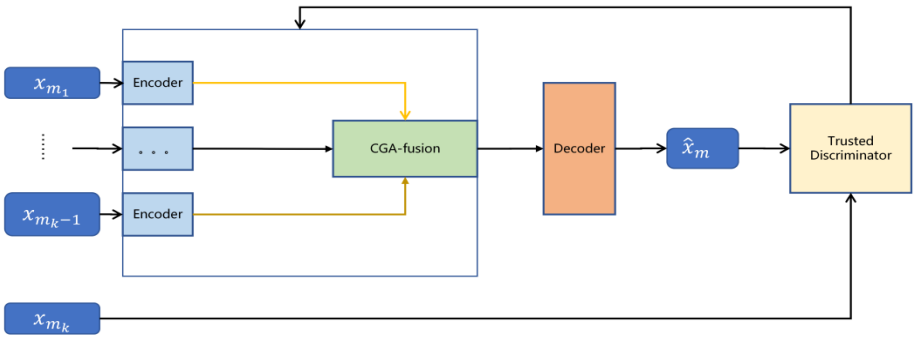


Figure 2. QHCSM -GAN.

3.2.1. Multi-channel feature fusion encoder

The encoder employs a K-way parallel autoencoder structure, comprising K-1 feature fusion branches and 1 guidance branch. Each branch extracts high-level semantic features and low-level texture detail features from the image through multi-layer coding modules, avoiding feature aliasing and detail loss. The network embeds a content-guided attention fusion module (CGA-fusion) that adaptively fuses cross-modal and multi-scale features, weighting and enhancing the effective features of cattle and sheep targets while suppressing interference from complex background noise in the plateau region (See **Figure 3**).

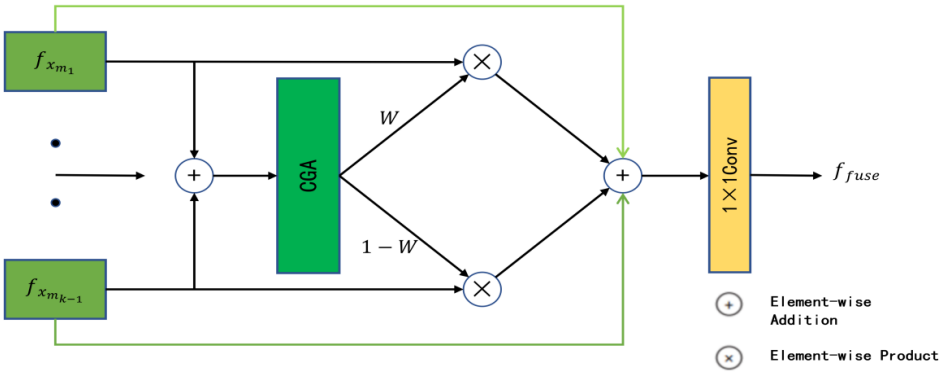


Figure 3. CGA -fusion module.

The guidance branch takes a real, clear image as input, extracts high-quality reference features, and performs deep supervision on the fusion encoder through feature alignment loss. The feature alignment loss is calculated as follows:

$$L_{feat} = \sum_{k\in\{h,l\}} \| F^k - \tilde{F}^k \|_1 \tag{3-1}$$

In the formula : F^k represents the features extracted by the multi-path feature fusion encoder, $\tilde{F}^k$ represents the ground truth features extracted by the guide autoencoder, $\| \|_1$ and represents the L1 paradigm .

The training of the entire multimodal conversion network is constrained by multiple loss mechanisms,including image-level reconstruction loss L_{img}, reconstruction loss for each modality feature L_{rec}, and the aforementioned feature alignment loss λ_{rec}. The total loss is defined as:

$$L_{MIN}=L_{img}+\lambda_{rec}L_{rec}+\lambda_{feat}L_{feat} \tag{3-2}$$

Among them , λ_{rec} and λ_{feat} are the balance coefficients.

The CGA fusion encoder effectively suppresses feature shifts caused by uneven lighting and image blurring in high-altitude areas through adaptive high- and low-frequency feature fusion; it enhances the stability of cattle and sheep target features through content-aware attention, mitigating the impact of posture and scale changes; and it weakens complex background interference through global attention, achieving early enhancement of foreground targets and background denoising, providing robust and clean feature representations for subsequent reconstruction.

3.2.2. Evidence theory credibility checker

Traditional generative adversarial networks (GANs) typically use Softmax to output classification confidence scores for their discriminators, which can easily lead to misclassifications due to “overconfidence” in the generated images. To address this issue and quantify the uncertainty in the generated results, this study employs a reliable classifier based on evidence theory as the discriminator D.

Evidence theory is an important mathematical tool for modeling and fusing uncertain information. It can quantify the conflict of image features, noise interference, and information loss through basic probability assignment, confidence function, and likelihood function. In the cattle and sheep image completion task, this

theory can effectively model the uncertainty caused by texture blur, occlusion, and illumination distortion, and improve the reliability of the discriminator in distinguishing between real and fake images.

The Dirichlet distribution, as the conjugate prior of a multinomial distribution, is often used for distribution modeling of probability vectors and classification confidence. Introducing it into a confidence discriminator can constrain the probability distribution of the evidence vector output by the network, calibrate the confidence level of the generated image, suppress unreasonable generation and artifacts, and make the completion result more consistent with the real distribution of the plateau cattle and sheep target in terms of structure, texture, and semantics.

This discriminator treats true/false binary classification as a subjective logical problem. For the input image, it doesn't directly output classification probabilities, but instead predicts a set of evidence vectors e to parameterize a Dirichlet distribution $Dir(p \mid \alpha)$. Here, the concentration parameter is denoted by $\acute{a} = e + 1$, and the class probability is represented by a simplex. Through this distribution, both the quality of classification belief p b_k and the overall uncertainty u can be obtained simultaneously.

$$u = \frac{S}{S+1}, b = \frac{e_k}{S+1} \tag{3-3}$$

Where $S = \sum_{k=1}^{K} (e_k + 1)$ k is the number of categories .

The discriminator's training loss L_D consists of two parts: first, an ensemble cross-entropy loss based on the Dirichlet distribution L_{ECE}, designed to gather more evidence for the correct class; and second, a KL divergence loss L_{KL}, used to penalize evidence for incorrect classes, preventing it from approaching zero.

$L_D = L_{ECE} + \lambda_{KL} L_{KL}$ The balancing factor λ_{KL} is gradually increased during training to ensure the network fully explores the parameter space. Correspondingly, in addition to the adversarial constraints from D L_{adv}, the generator G is also subject to an additional evidence loss L_{evi}. This loss ensures that the images produced by the generator not only closely approximate reality in terms of discrimination, but also that the evidence contained in their internal features supports the " real " category, thereby achieving feature-level alignment.

$$L_{\text{evi}} = \left\| e_{fake} - e_{real} \right\|_2^2 \tag{3-5}$$

Finally, the overall loss function of the generator is:

$$L_G = L_{MIN} + \lambda_{\text{adv}} L_{adv} + \lambda_{evi} L_{evi} \tag{3-6}$$

By leveraging the dual constraints of the trusted discriminator (adversarial loss and evidence loss), this method can effectively quantify and utilize uncertainty during image generation, thereby synthesizing structurally accurate, detailed, and reliable missing modal images.

3.3. Improved YOLO-OC detection model design

This paper uses the lightweight YOLOv8n as the baseline model. To address the problems of difficulty in adapting to multi-scale targets and strong background interference in plateau scenes, the backbone network and the neck network are optimized respectively to construct the YOLO-OC detection model. The feature extraction capability is optimized by using ODConv full-dimensional dynamic convolution, and the key feature focusing is enhanced by combining the CBAM attention mechanism. **Figure 4** shows the complete

network structure of YOLO-OC, with the embedding positions of ODConv and CBAM and the overall network flow marked.

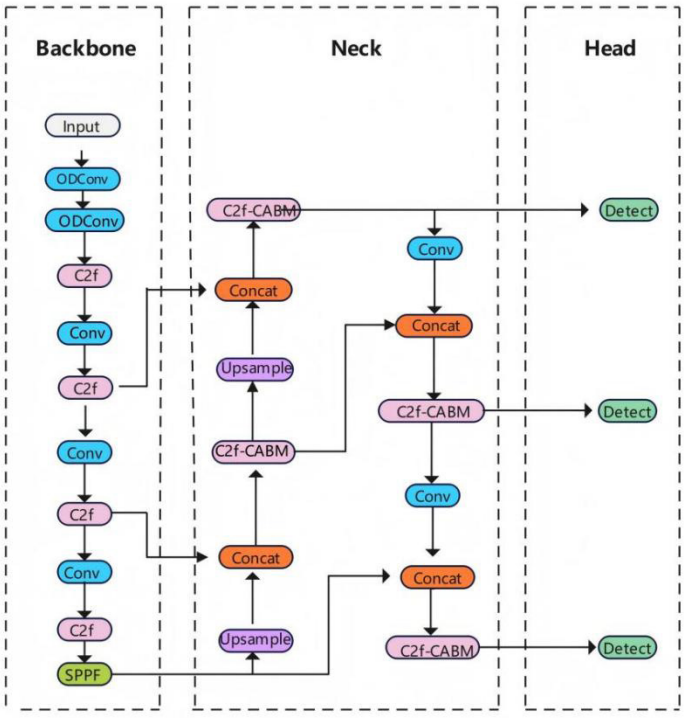


Figure 4. Optimization structure diagram of YOLO-OC model.

3.3.1. Backbone network optimization based on ODConv

Traditional fixed convolutional kernels can only extract features at a single scale, failing to adapt to the scale variations of large near-field targets and small distant targets such as cattle and sheep in the plateau region. This paper replaces the shallow fixed convolutions in the backbone network with full-dimensional dynamic convolutions (ODConv). Through a multi-dimensional attention mechanism involving space, channels, and convolutional kernels, convolutional weights are dynamically generated. This allows for adaptive adjustment of the response mode based on the input image features, effectively enhancing the extraction of detailed features from small targets and the aggregation of semantic features from large targets, thus solving the problem of insufficient scale adaptability of fixed convolutions (See **Figure 5**).

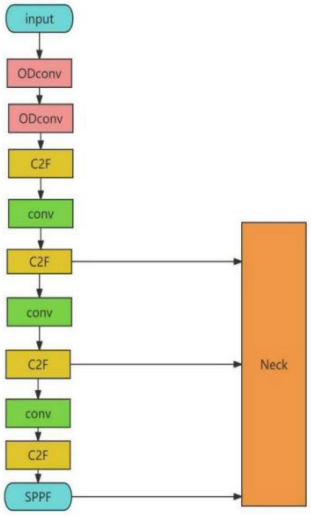


Figure 5. Structure diagram of ODC-FPN network.

ODConv generates dynamic weights through an attention mechanism, enabling adaptive adjustment of the convolutional kernel. The formula for generating dynamic weights is as follows:

$$W_{\mathrm{d}ynamic} = \sum_{i=1}^{N} \alpha_i \cdot W_i \tag{4-1}$$

Where is $W_{dynamic}$ the generated dynamic convolution kernel weight, is the number of basic weight vectors, α_{i} is the dynamic weight coefficient generated by the attention mechanism, W_i and is the i-th basic weight vector .

$$\sum_{i=1}^{N} \alpha_{\mathrm{i}} = 1 \tag{4-2}$$

Through the above mechanism, ODConv can dynamically adjust the convolution kernel parameters based on the input features, thereby improving the model's adaptability to targets of different scales.

3.3.2. CBAM-based neck network optimization

To address the challenges presented by the QH-Livestock dataset, such as complex backgrounds in high-altitude pasture scenes, livestock coat colors closely resembling environmental textures, and the tendency for target features to weaken under strong light, this chapter optimizes the feature fusion process in the Neck section of YOLO v8 by embedding Convolutional Block Attention Modules (CBAMs) before and after the feature concatenation positions within the C2f module. By introducing a joint modeling mechanism of channel attention and spatial attention, the model's responsiveness to key regions and discriminative features of cattle and sheep is enhanced, while suppressing irrelevant interference from complex backgrounds and lighting variations. This improves the model's feature focusing ability and detection stability in high-altitude scenes.

CBAM is a lightweight attention mechanism module. Its core idea is to adaptively model the importance of different channels and spatial locations in the feature map through a concatenated structure of channel attention and spatial attention, and dynamically adjust the feature response weights to highlight target-related information, suppress background noise interference, and improve the model's feature representation ability in complex scenes. Compared with methods that only focus on single-dimensional feature enhancement,

CBAM can optimize feature representation from both “channel selection” and “spatial localization” levels, making it more suitable for feature enhancement in target detection tasks in complex environments.

CBAM first models the importance of the input feature map in the channel dimension through the Channel Attention Module (CAM). This module performs global average pooling and global max pooling on the input features, and then maps the resulting features into a shared multilayer perceptron to generate channel attention weights. The calculation process can be represented as follows:

$$M_C(F) = \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F))) \tag{4-3}$$

Where $\sigma()$ is the Sigmoid activation function, $\mathrm{MLP}()$ is the multilayer perceptron, is the global average pooling, and is the global max pooling .

After channel-weighting, CBAM further models the important regions of the feature map in the spatial dimension through the Spatial Attention Module (SAM). This module first performs average pooling and max pooling operations on the channel-weighted feature map, then concatenates the results and feeds them into a convolutional layer to generate the spatial attention weight map. The calculation process can be represented as follows:

$$M_S(F) = \sigma(Conv([AvgPool(F); MaxPool(F)])) \tag{4-4}$$

Conv This represents a convolution operation, [;]indicating feature map concatenation.

Embedding this module into the feature fusion process of the Neck part helps to improve the effectiveness of multi-scale features in the transmission and fusion stages, thereby enhancing the model’s ability to distinguish targets under complex backgrounds, lighting changes, and local occlusion conditions.

As shown in **Figure 6**, by introducing CBAM into the Neck part, the model can further enhance its attention to the target region while maintaining the basic stability of the original network framework, providing more discriminative fusion features for subsequent detection heads, thereby improving the false detection and false negative detection problems in plateau scenes.

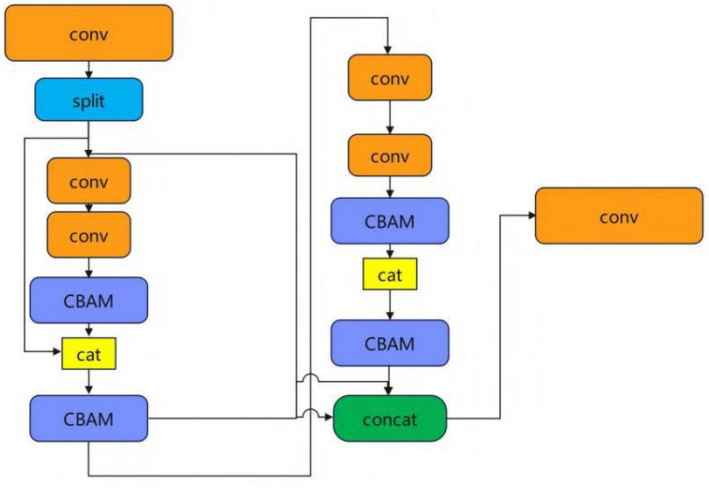


Figure 6. Network structure diagram of C2F-CBAM-Neck.

4.Experiment and results analysis

4.1. Experimental setup

All experiments in this paper were conducted on an Ubuntu 22.04 LTS system, using the PyTorch 2.1.2 deep learning framework and CUDA 11.6 parallel computing environment. The hardware configuration consisted of an Intel i7-8750H CPU, 32GB of RAM, and an NVIDIA RTX 4060Ti GPU. The experiments were conducted on the QH-Livestock dataset, with 37% of the training samples enhanced and optimized using the QHCSM-GAN algorithm. The model training parameters were uniformly set as follows: input image size 640 × 640, SGD optimizer, initial learning rate 0.01, batch size 8, momentum coefficient 0.937, and a maximum of 300 training epochs.

For the image completion task, root mean square error (RMSE), structural similarity (SSIM), peak signal-to-noise ratio (PSNR), and learned perceptual patch similarity (LPIPS) are selected as evaluation metrics; for the object detection task, precision (P), recall (R), mAP50, and inference speed (FPS) are selected to comprehensively evaluate the model's accuracy and real-time performance.

4.2. Experimental results of QHCSM-GAN image completion

To verify the superiority of the QHCSM-GAN algorithm, eight mainstream image completion algorithms, including StarGAN, Pix2Pix, and TokenFusion, were selected for comparative experiments. The tests were completed on the QH-Livestock, BraTS2020, and RaFD datasets [8–14].

Table 1 presents the quantitative comparison results of the QH-Livestock dataset. In the visible light to infrared modal restoration task, QHCSM-GAN achieves an SSIM of 0.9345, a PSNR of 28.92 dB, and an LPIPS as low as 0.0825, significantly outperforming the comparison algorithms in all metrics. In the infrared-to-visible light restoration task, QHCSM-GAN achieves SSIM and PSNR of 0.8983 and 27.14 dB, respectively, demonstrating the best overall performance.

Table 1. Comparison of modal synthesis performance of various methods on the QH-Livestock dataset

Methods	Visible to Infrared			Infrared to Visible		
	SSIM↑	PSNR↑(dB)	LPIPS↓	SSIM↑	PSNR↑(dB)	LPIPS↓
StarGAN	0.8365	22.48	0.1615	0.8395	22.12	0.1735
CollaGAN	0.7887	21.43	0.2272	0.8091	22.88	0.1892
Hi-net	0.8521	23.15	0.1426	0.8478	23.02	0.1568
Pix2Pix	0.8798	24.89	0.1087	0.8645	24.83	0.1245
AttentionGAN	0.8568	23.37	0.1086	0.8231	21.23	0.1322
TSIT	0.8335	22.46	0.1625	0.8392	22.25	0.1735
CUT	0.8172	21.18	0.1334	0.8134	20.52	0.1591
TokenFusion	0.9011	25.67	0.0932	0.8435	23.58	0.1396
Ours(QHCSM-GAN)	0.9345	28.92	0.0825	0.8983	27.14	0.1160

Table 2 presents the ablation experiment results of QHCSM-GAN. Feature alignment loss and evidence loss are crucial to model performance; removing any constraint term leads to a significant decrease in image restoration accuracy, demonstrating that the dual constraint mechanism effectively ensures the structural consistency and generation reliability of the restored image. Cross-dataset experiments further validate the algorithm's generalization ability.

Table 2. Results of the QHCSM-GAN ablation experiment

Model configurations	SSIM↑	PSNR↑(dB)	LPIPS↓
QHCSM-GAN - feat -evi	0.9 34 5	28.92	0.08 5 5
QHCSM-GAN - feat	0.9113	26.81	0.1011
QHCSM-GAN - evi	0.9064	26.53	0.1022

Figure 7 shows the visual comparison results of image completion by various algorithms. The algorithm presented in this paper can effectively repair defects such as occlusion and overexposure in plateau images, restore the complete outline and texture details of cattle and sheep, and provide high-quality training samples for the detection model.

Figure 7. The image on the left is the original image; the image on the right is the one generated by QHCSM-GAN.

4.3. Results of YOLO-OC comparison experiment

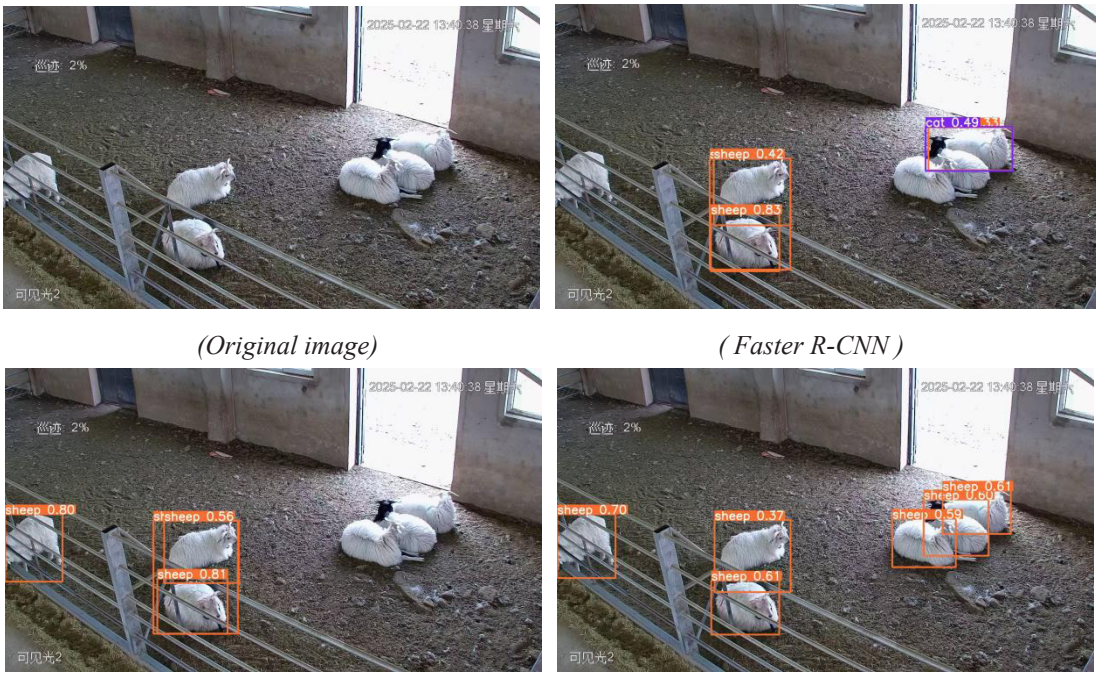
To verify the detection performance of the YOLO-OC model, a comparative experiment was conducted with seven mainstream detection models, including Faster R-CNN, YOLOv5n, YOLOv7-tiny, YOLOv8n, YOLOv9-tiny, and YOLOv12 [15].

Table 3 presents the quantitative comparison results of the models. The YOLO-OC model in this paper achieved a precision of 84.6%, a recall of 81.2%, and a mAP50 of 82.5%, representing improvements of 6.1%, 8.9%, and 7.3%, respectively, compared to the baseline YOLOv8n. Compared to the latest lightweight YOLO model, YOLO-OC achieves a significant improvement in accuracy with minimal speed loss (10^9 FPS), fully meeting the real-time monitoring needs of high-altitude pastures. Compared to the two-stage Faster R-CNN model, it has an absolute advantage in both accuracy and speed.

Table 3. Performance comparison of each model on the dataset

Model name	Accuracy	Recall rate	mAP50	Fps	Params
Faster R-CNN	72.3 %	58.6 %	65.4 %	15	41.8M
YOLO v5 n	74.5 %	68.2 %	71.3 %	84	**1.9M**
YOLO v7-tiny	76.8 %	71.5 %	74.2 %	87	6.2M
YOLO v8 (benchmark)	78.5 %	72.3 %	75.2 %	90	3.2M
YOLO v 9 -tiny	79.2%	74.8%	77.5%	115	2.0M
YOLO v 12	0.81.4%	76.5%	78.9%	**120**	2.6M
YOLO-OC (Optimized)	**84.6 %**	**81.2 %**	**82.5 %**	109	3.3M

Figure 8 shows the visualization comparison results of different detection models. Mainstream algorithms have serious problems with missed detections and false detections in small targets, strong lighting, and complex background scenes, while YOLO-OC can accurately identify cattle and sheep targets at multiple scales, and its anti-interference ability and scene robustness are significantly better than the comparison models.



(Original image) *(Faster R-CNN)*

(YOLO v8) *(YOLO -OC)*

Figure 8. Comparative image visualization results.

4.4. Results of YOLO-OC ablation experiments

To verify the effectiveness and synergistic effect of the ODConv and CBAM modules, an ablation experiment was set up.

Table 4 presents the quantitative results of the ablation experiments. When only the ODConv module is introduced, the model's mAP50 is improved by 2.6%, demonstrating that dynamic convolution can effectively enhance multi-scale feature extraction capabilities. When only the CBAM module is introduced, the mAP50 is improved by 3.2%, verifying that the attention mechanism can efficiently suppress background noise. The model performance reaches its optimal level after the fusion of the two modules, forming a collaborative optimization mechanism of "dynamic feature extraction + precise feature focusing".

Table 4. Results of the ablation experiment

Model configuration	**Accuracy(P)**	**Recall rat**e(R)	**mAP50**	**Reasoning** speed(FPS)
YOLO v8 (benchmark)	78.5 %	72.3 %	75.2 %	90
YOLO v8 - ODConv	80.3 %	75.6 %	77.8 %	85
YOLO v8 - CBAM	81.7 %	74. 2 %	78.4 %	86
YOLO-OC	84.6 %	81.2 %	82.5 %	109

Figure 9 shows the visual comparison results of the ablation experiment, further demonstrating that the dual-module improvement can significantly enhance the detection effect in challenging high-altitude scenarios.

(YOLO-oc) *(YOLOv8)*

Figure 9. Ablation image visualization results.

4.5. Training process analysis

Figure 10 shows the training loss and accuracy curves of the YOLO-OC model. The model converges quickly and is highly stable, with all accuracy metrics rising rapidly within 100 rounds and stabilizing after 200 rounds, showing no significant oscillations or overfitting. Thanks to the high-quality data augmentation and refined module optimization of QHCSM-GAN, the model possesses excellent generalization ability and stability for engineering deployment.

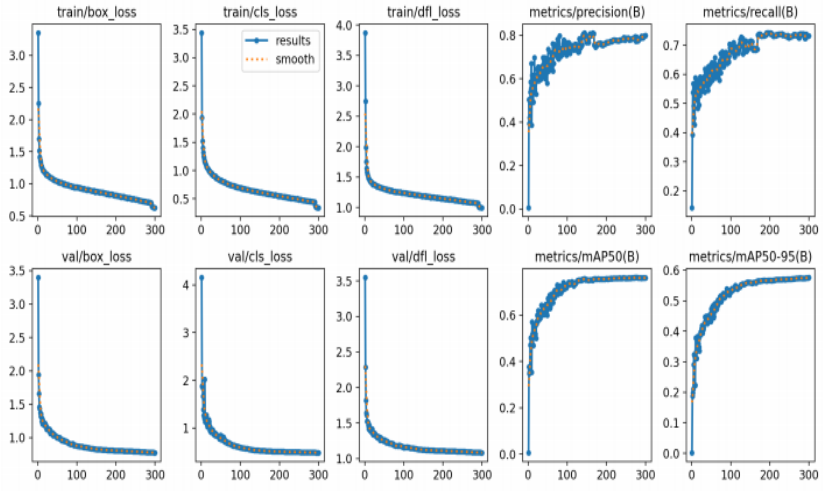


Figure 10. Performance trend chart.

5.Conclusions and outlook

To address the industry pain points in intelligent monitoring of livestock in high-altitude areas, such as poor image quality, scarcity of dedicated datasets, and low detection accuracy in complex scenes, this paper constructs an intelligent detection framework for cattle and sheep in high-altitude areas, integrating dataset

construction, image completion optimization, and target detection improvement. First, the QH-Livestock dataset, specifically designed for cattle and sheep in high-altitude areas, is constructed, filling the data resource gap in high-altitude grazing scenarios. Second, the QHCSM-GAN image completion algorithm is proposed, achieving high-fidelity restoration of degraded images in high-altitude areas through multi-path feature fusion and uncertainty constraints based on evidence theory, effectively improving the quality of training samples. Finally, a YOLO-OC detection model is designed based on YOLOv8n, integrating ODConv dynamic convolution and CBAM attention mechanisms to significantly improve the model's multi-scale detection capability and resistance to interference in complex backgrounds. Multiple experimental results demonstrate that the proposed framework exhibits high accuracy, good real-time performance, and strong robustness, effectively adapting to intelligent monitoring scenarios in open pastures in high-altitude areas.

This study still has certain limitations: the dataset does not adequately cover extreme rain and snow weather and low-light nighttime scenarios; the model's lightweight design still has room for improvement, and its ability to adapt to low-power edge deployments needs to be enhanced. Future research will focus on three aspects: first, expanding the dataset to enrich samples for challenging scenarios such as extreme weather and nighttime monitoring; second, further lightweighting the model structure to reduce the number of parameters and computational load, making it suitable for embedded edge device deployments; and third, integrating target tracking and behavior recognition algorithms to construct a comprehensive and multifunctional intelligent monitoring system for plateau animal husbandry, thereby facilitating the intelligent and digital transformation and upgrading of plateau animal husbandry.

Disclosure statement

The authors declare no conflict of interest.

References

[1] Goodfellow I, Pouget-Abadie J, Mirza M, et al., 2014, Generative Adversarial Nets. Proceedings of the 28th International Conference on Neural Information Processing Systems, 2672–2680.

[2] Tian Y, Ye Q, Doermann D, 2025, YOLOv12: Attention-Centric Real-Time Object Detectors. arXiv.

[3] Li C, Zhou A, Yao A, 2022, Omni-Dimensional Dynamic Convolution. arXiv:2209.07947.

[4] Chen M, Zhou G, Ma Y, 2024, An Improved YOLOv8 Small Target Detection Algorithm With Embedded CBAM Attention. Computer Engineering and Science, 46(03): 478–485.

[5] Wang Y, Xu D, Wang Y, et al., 2025, TAF-ELAN: Tri-Attention Fusion Enhanced ELAN Network for Single Image Super-Resolution. Proceedings of the 2025 7th International Academic Exchange Conference on Science and Technology Innovation, 163–167.

[6] Liu S, Qi L, Qin H, et al., 2018, Path Aggregation Network for Instance Segmentation. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 8759–8768.

[7] Ma M, Liang L, Sheng X, 2025, MDF-Net: An Attention-Guided Multi-Scale Dual-Fusion Network for Retinal Vessel Segmentation. Measurement, 257(02): 118695.

[8] Choi Y, Choi M, Kim M, et al., 2018, StarGAN: Unified Generative Adversarial Networks for Multi-Domain Image-to-Image Translation. Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 8789–8797.

[9] Isola P, Zhu J, Zhou T, et al., 2017, Image-to-Image Translation with Conditional Adversarial Networks. Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition, 5967–5976.

[10] Tang H, Xu K, Li X, et al., 2019, AttentionGAN: Unpaired Image-to-Image Translation Using Attention-Guided Generative Adversarial Networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 5647–5656.

[11] Jiang L, Li M, Lin L, et al., 2021, TSIT: A Simple and Versatile Framework for Image-to-Image Translation. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 423–432.

[12] Park T, Efros A, Zhang R, et al., 2020, Contrastive Learning for Unpaired Image-to-Image Translation. European Conference on Computer Vision, 319–334.

[13] Kim J, Kim M, Kim D, et al., 2019, CollaGAN: Collaborative Generative Adversarial Network for Missing Image Data Imputation. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 248–256.

[14] Zhou T, Li Z, Wang H, et al., 2020, Hi-Net: Hybrid-Fusion Network for Multi-Modal MR Image Synthesis. IEEE Transactions on Medical Imaging, 39(12): 4014–4025.

[15] Girshick R, 2015, Fast R-CNN. Proceedings of the IEEE International Conference on Computer Vision, 1440–1448.