JEI Journal of Electronic Imaging

RESEARCH PAPER

UD-YOLOv5s: Recognition of cattle regurgitation behavior based on upper and lower jaw skeleton feature extraction

Guohong Gao, Chengchao Wang, Jianping Wang,* Yingying Lv, Qian Li, Xueyan Zhang, Zhiyu Li, and Guanglan Chen

Henan Institute of Science and Technology, School of Information Engineering, Xinxiang, China

ABSTRACT. Rumination plays a pivotal role in assessing the health status of ruminants. However, conventional contact devices such as ear tags and pressure sensors raise animal welfare concerns during rumination behavior detection. Deep learning offers a promising solution for non-contact rumination recognition by training neural networks on datasets. We introduce UD-YOLOv5s, an approach for bovine rumination recognition that incorporates jaw skeleton feature extraction techniques. Initially, a skeleton feature extraction method is proposed for the upper and lower jaws, employing skeleton heatmap descriptors and the Kalman filter algorithm. Subsequently, the UD-YOLOv5s method is developed for rumination recognition. To optimize the UD-YOLOv5s model, the traditional intersection over the union loss function is replaced with the generalized one. A self-built bovine rumination dataset is used to compare the performance of three deep learning techniques: mean shift algorithm, mask region-based convolutional neural network, and you only look once version 3 (YOLOv3). The results of the ablation experiment demonstrate that UD-YOLOv5s achieves impressive precision (98.25%), recall (97.75%), and a mean average precision of 93.43%. We conducted a generalization performance evaluation in a controlled experimental environment to ensure fairness, indicating that UD-YOLOv5s converges faster than other models while maintaining comparable recognition accuracy. Moreover, our work reveals that when convergence speed is equal, UD-YOLOv5s outperforms other models regarding recognition accuracy. These findings provide robust support for accurately identifying cattle rumination behavior, showcasing the potential of the UD-YOLOv5s method in advancing ruminant health assessment.

Keywords: ruminant; behavior recognition; YOLOV5s; skeleton extraction

Paper 230539G received May 5, 2023; revised Jul. 22, 2023; accepted Aug. 11, 2023; published Aug. 29, 2023.

1 Introduction

Ruminating is a distinctive digestive process observed in cattle, characterized by repetitive chewing, swallowing, and regurgitation, typically in peaceful surroundings such as grazing fields.[1] The capability to recognize and track ruminating behavior in cattle presents an opportunity for enhancing livestock health management within the industry. Swiftly detecting and addressing any anomalies in cattle behavior can help mitigate potential health problems. Currently, diverse methodologies are employed to identify ruminating behavior in cattle, encompassing sensor technology, computer vision, behavior classification, and deep learning techniques.[2]

*Address all correspondence to Jianping Wang, wangjianping@hist.edu.cn

Using sensor technology is one of the most commonly employed methods for identifying rumination behavior in cattle. This approach involves detecting biological signals produced during rumination, such as changes in electrical potential resulting from chewing movements and sounds generated by gastrointestinal motility.[3] While sensors offer reliable data, they require physical contact with animals and may raise concerns regarding animal welfare. Another frequently used technique is computer vision, which entails recording visual information using cameras and analyzing and processing video frames to extract rumination features. For example, contour extraction technology can be utilized to obtain cattle outlines.[4] However, it is essential to consider the changing rumination scenes and lighting conditions when utilizing this method. Behavior classification offers a promising means of identifying rumination behavior using machine learning techniques. This approach involves training a classifier to recognize rumination behavior automatically. Supervised learning algorithms such as *k*-nearest neighbors, support vector machines, random forests, etc., can be employed to train the data, and the classifier's performance can then be evaluated using test data. However, ensuring accurate results is crucial, as this method requires a precise dataset and significant manual labeling work.

Deep learning skeleton extraction is an innovative technique that leverages deep learning methods to extract object skeletons from images. This process involves two main steps: object contour detection and skeleton extraction. Initially, deep learning algorithms are employed to detect the object's contour line, which is the foundation for extracting its skeleton, considering its shape and spatial relationships within the image. Subsequently, data about position, posture, and state can be determined by analyzing the skeleton in the image. This approach significantly enhances the accuracy of machine vision systems. The most commonly utilized techniques for deep learning skeleton extraction include convolutional neural networks (CNN), deep neural networks (DNNs), and recurrent neural networks.[5] During training, input pixel information undergoes multiple layers of convolution and pooling operations, creating high-level feature representations. For a given image, the network utilizes feedback to extract more stable feature representations, allowing for the construction of different types of deep learning models tailored to the specific requirements of the target task.

You only look once version 5 (YOLOv5) is an advanced deep learning model designed for precise object detection, efficiently identifying object locations and classes in images with remarkable accuracy.[6] YOLOv5 comprises four basic models: YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x. Among these, YOLOv5s stands out with its compact architecture and narrow feature width, making it particularly suitable for applications requiring high detection accuracy. As an evolution of single-stage object detection models, YOLOv5s significantly enhances network performance, achieving superior algorithmic accuracy and speed. This improvement is achieved by introducing additional layers and convolutional kernels to improve feature extraction and understanding while utilizing the cross-entropy loss function to refine predictions for different targets. YOLOv5s has gained widespread adoption across various domains, including intelligent surveillance, smart agriculture, and automated monitoring, due to its exceptional accuracy, rapid processing capabilities, and real-time detection and tracking of multiple objects.[7]

This paper introduces upper and lower you only look once version 5 small (UD-YOLOv5s), an innovative method for recognizing cattle rumination behavior based on extracting upper and lower jaw skeleton features using the YOLOv5s network. This work is the first to combine the YOLOv5s network with upper and lower jaw skeleton features for cattle rumination behavior recognition. The key contributions of this research are as follows:

1. A novel feature extraction method was developed for cattle's upper and lower jaw skeleton, employing skeleton heatmap descriptors (SHDs) and Kalman filter (KF) algorithms.
2. We propose a cattle rumination recognition method named UD-YOLOv5s, which utilizes YOLOv5s as the backbone network.
3. The performance of UD-YOLOv5s was further enhanced by replacing the traditional intersection over union (IoU) loss function with the generalized IoU (GIoU) loss function.
4. The effectiveness of the UD-YOLOv5s network was evaluated using a self-built dataset, and model validation was conducted through comparative experiments with other algorithms, including mean shift algorithm (MEAN-SHIFT), mask region-based convolutional neural network (MASK-RCNN), and you only look once version 3 (YOLOv3).

The rest of this paper is organized as follows: Sec. 2 presents an overview of the current state of research in this field. Section 3 details the basic architecture of UD-YOLOv5s and the related processing methods. Section 4 outlines the materials and techniques. An analysis of the experimental results is provided in Sec. 5. Finally, in Sec. 6, we draw conclusions based on the findings from this research.

2 Research Status

2.1 Progress in Ruminant Identification Technology

Borchers et al.[8] conducted a comprehensive work on rumination behavior in cattle, employing six different tri-axial accelerometer technologies. These included the cattle manager sensor, used to monitor rumination and feeding time, and the intelligent bow sensor, used to track cattle and monitor their rumination behavior and time spent in feeding areas. Visual observation was utilized to generate behavior time lengths, recording the start and end times of rumination behaviors occurring within a day. In another work, Bishop-Hurley et al.[9] employed a collar system equipped with tri-axial accelerometers and magnetometers to investigate feed additives' impact on cattle feeding behavior. In addition, they analyzed and developed behavior models using multivariate time series data to address potential issues with the sensors. Arablouei et al.[10] developed a specialized pipeline for pre-processing, feature extraction, and cattle behavior classification using measurement data. This pipeline was designed to address resource constraints. They collected data from 10 cattle using tri-axial accelerometer sensors on collar tags to record their rumination and other behaviors. Watt et al.[11] utilized acoustic methods to measure rumination activity in cows and discovered a positive correlation between dry matter intake and rumination time. Furthermore, they conducted a 10-day baseline detection for frequency analysis of rumination time.

Rombach et al.[12] validated the effectiveness of the ruminant welfare system (RWS) in measuring cows' rumination behavior during barn grazing and supplementation periods. They improved the algorithm used in the RWS system. They found that the two converters used in the experiment were better at distinguishing rumination and feeding behavior in cows during other activities. Braun et al.[13] established feature pressure curves by comparing data obtained from pressure sensors with 24-h direct observation data, demonstrating the regularity of chewing during rumination and generating a uniformly regular waveform. The consistency of the results was confirmed by comparing the data obtained through direct observation and pressure sensors. Gregorini et al.[14] conducted a 21-day strip grazing experiment on a perennial ryegrass pasture with 8 cows equipped with high resolution (HR) tags and rumination collars to record rumination behavior and chewing activity during rumination. They calculated and registered rumination time based on the mean interval time of chewing actions, ultimately verifying the cows' daily rumination pattern under the constraints of the pasture's time. Handcock et al.[15] demonstrated the potential of a wireless sensor network to monitor cattle behavior by combining global positioning system collars with satellite imagery, providing high temporal resolution. Schirmann et al.[16] evaluated changes in rumination and feeding behavior before and after calving using time standards. Zehner et al.[17] developed a novel scientific detection device to automatically measure stable-fed cow rumination and feeding behavior using a universal algorithm based on animal-specific learning data, with two software versions for the system. Pereira et al.[18] evaluated the correlation and differences between direct visual observation and sensor data analysis by recording data for 6 h using a trained observer and ear-tag accelerometer sensors. Ruuska et al.[19] proposed a pressure sensor system to measure rumination time in cattle and validated its effectiveness through comparative experiments.

Porto et al.[20] proposed a computer vision-based automated detection system for free-stall resting behavior in cattle that utilizes a Viola–Jones algorithm for cow lying behavior detection. The system captures image data of the research area through a multi-camera recording system and validates the system's effectiveness by comparing the detection results with visual recognition results. Yazdanbakhsh et al.[21] developed an intelligent system for continuous monitoring of the health status of each animal in livestock using sensors installed on animals. The experiment showed that higher sensitivity and specificity could be achieved using an integrated classifier in the wavelet domain. Arcidiacono et al.[22] proposed a model based on an acceleration thresholding

algorithm to detect cattle behavior by predicting the threshold of behavior onset from data obtained under specific experiments in the barn where the cattle is housed, thus determining the behavior of the cattle. Viazzi et al.[23] developed a movement model based on lame cow walking behavior, increasing both single threshold accuracy and actual positive rate by representing the behavioral features hierarchically, effectively improving the accuracy of rapid and precise identification of lame cow walking on the farm. Pahl et al.[24] analyzed feeding characteristics recorded by weighing troughs and rumination time changes recorded by acoustic sensors based on the feeding characteristics before and after cow estrus. Lee et al.[25] organized and analyzed the predicted physiological parameters and various data from a wearable wireless sensor system for cows. The sensors used in the system primarily focus on rumination behavior, including ear tags, collars, and reticulum boluses. Jonsson et al.[26] introduced individual animal lying balance by combining information from step sensors and leg tilt sensors, deriving a new change detection scheme, and establishing a change detection algorithm to describe cow behavior state within a given time interval using a binary variable. Braun et al.[27] employed an automated system to assess feeding and rumination variables in a cohort of 300 dairy cows over 24 h. The method utilized pressure sensors integrated into neck collars to capture jaw movements. Subsequently, it analyzed vital parameters such as the duration of rumination, number of chewing cycles, number of chews per cycle, and units of chews per cycle. These studies have contributed significantly to understanding cattle rumination behavior and have used various sensor technologies and data analysis techniques to achieve meaningful insights.

2.2 Advances in Deep Learning-Based Cattle Regurgitation Recognition Technology

Jiang et al.[28] proposed implementing the fast layered YOLO version 3 (FLYOLOv3) network for detecting essential parts of cattle in complex scenes based on the filter layer. They integrated a customized filter layer with an average filtering algorithm and leaky rectified linear unit (leaky ReLU) function to reduce training interference. Shakeel et al.[29] introduced an innovative behavior recognition and computing scheme to predict cattle behavior. The proposed method used a deep recurrent learning paradigm to cycle the recognition pattern and classify abnormal situations based on differentiated data patterns. Chen et al.[30] evaluated the latest developments in computer vision methods for recognizing cattle behavior based on animal productivity, health, and welfare. They analyzed the effect of image segmentation, recognition, and behavior recognition using traditional computer vision and deep learning methods, listing the progress of crucial research in this field. Peng et al.[31] combined long short-term memory (LSTM) networks to detect and identify cattle behavior using inertial measurement units and classify behaviors such as regurgitation and feeding. LSTM - recurrent neural network (LSTM-RNN) model training showed potential for development in cattle regurgitation detection. Tamura et al.[32] investigated the correlation between cattle behavior and acceleration data collected using a tri-axis neck-mounted accelerometer. They proposed the feasibility of improving behavior classification accuracy through machine learning. By recording the characteristic acceleration waves of eating, rumination, and lying down during visual observation of cattle behavior, they combined the farm dataset for decision-tree learning and ultimately verified the accuracy of decision-tree knowledge.

Fuentes et al.[33] proposed a hierarchical cattle behavior recognition method based on deep learning, incorporating spatiotemporal information. The framework includes appearance features at the frame level and spatiotemporal information containing more contextual time features. The designed system detected and located rumination behavior in multiple cattle in video frame regions. The method was validated using datasets captured in both day and night environments, showing effective recognition of 15 types of hierarchical activities. McDonagh et al.[34] continuously monitored 46 cattle and used image recognition technology to predict their behaviors. Non-local neural networks were trained and validated on video clips for each behavior, showing successful recognition and classification of behaviors over extended periods. Dutta et al.[35] classified cattle behavior recorded by a collar system equipped with a tri-axial accelerometer and magnetometer using machine learning techniques. A hybrid framework was developed to work the natural structure of sensor data, and various classification methods were compared to verify the model's superiority in representing rumination behavior. Chen et al.[36] proposed an intelligent monitoring method for cattle rumination behavior based on video analysis. They used the

MEAN-SHIFT algorithm to track the movement of the cattle's lower jaw and extract the centroid trajectory curve of the cattle's mouth movement from the video, enabling monitoring of cattle rumination behavior. Li et al.[37] proposed a multi-target monitoring method based on optical flow and frame difference for cattle rumination behavior. Using optical flow and frame difference, they identified candidate rumination cattle oral regions and tracked the cattle's mouth within the area, enabling automatic monitoring of the rumination cattle's mouth area. Tamura et al.[38] developed a method for detecting chewing speed using a tri-axial accelerometer connected to the cattle's neck. Using the fast Fourier transform algorithm, they analyzed neck vibration as rumination movement and calculated chewing time. Dutta et al.[39] proposed an intelligent IoT device for cattle monitoring to identify rumination behavior and collect rumination data through classification. The device was mounted on the cattle's neck, and data were transmitted to IoT servers and a cellular global-local space modulation module. Xu et al.[40] constructed an advanced instance segmentation framework MASK-RCNN to address the problem of cattle dataset occlusion and overlap in complex scenes. They compared classical algorithms and validated the best threshold and complete detection for cattle herd data in different situations, demonstrating the effectiveness of the proposed framework. These studies provide valuable insights into monitoring cattle behavior and have practical applications in the agricultural industry.[41–43]

2.3 Summary

The present work introduces an innovative technique for bovine rumination identification that addresses various perspectives. While rumination detection techniques based on wearable devices have limitations and may not meet animal welfare requirements, methods using pressure sensors and tri-axial accelerometers to detect rumination tend for prioritizing data collection over accurate identification. In addition, some approaches rely on identifying bovine body or head regions, which may need more accuracy in rumination recognition. In contrast, our proposed method focuses on bovine rumination identification by extracting features from the upper and lower jawbone skeleton. Specifically, we utilize skeleton extraction technology to extract the upper and lower jawbone skeleton features from the bovine mouth region. These features are then processed using the UD-YOLOv5s network architecture, enhancing the identification process's accuracy and efficiency. We optimize the loss function to improve performance. By adopting this method, we achieve more accurate and efficient bovine rumination identification compared to traditional techniques. It ensures higher precision and considers the animals' welfare, making it a promising step toward effective and humane bovine rumination identification.

3 UD-YOLOV5s-Based Method for Identifying Cattle Regurgitation Behavior

3.1 UD-YOLOv5s System Architecture

This paper introduces UD-YOLOv5s for accurately identifying cattle grazing behavior. The proposed approach leverages the YOLOv5s network to extract crucial skeletal features from upper and lower jawbones. We combine the SHD and KF algorithms to achieve this and process the original feature map. Subsequently, the focus block is utilized for image segmentation, and the output is unsampled and fed into the convolution layer. The feature fusion module combines these features, subsequently connected to the pre-processing layer. Finally, we optimize the output using the GIoU loss function. The overall architecture of our system is shown in Fig. 1. The ruminant behavior of cattle primarily manifests through the skeletal movements of their upper and lower jaw regions. In this paper, we extract the skeleton features of the upper and lower jaw regions separately to discriminate ruminant behavior accurately. Specifically, the upper jaw region exhibits linear movement during cow rumination. We employ the SHD algorithm to extract the upper jaw skeleton structure from the cow's image. The SHD algorithm is proficient in extracting the primary structural lines of objects by refining the edges of the image, thereby obtaining the skeleton features of the upper jaw. For the image of the cow's upper jaw, the SHD algorithm effectively simplifies the complex edge structure into skeleton lines, thereby extracting the main structural features of the upper jaw. In addition, we utilize the KF, a filtering algorithm for state estimation while extracting the cow's jaw skeleton features. The KF algorithm estimates the state of the jaw based on observed jaw position and motion trajectory, subsequently extracting

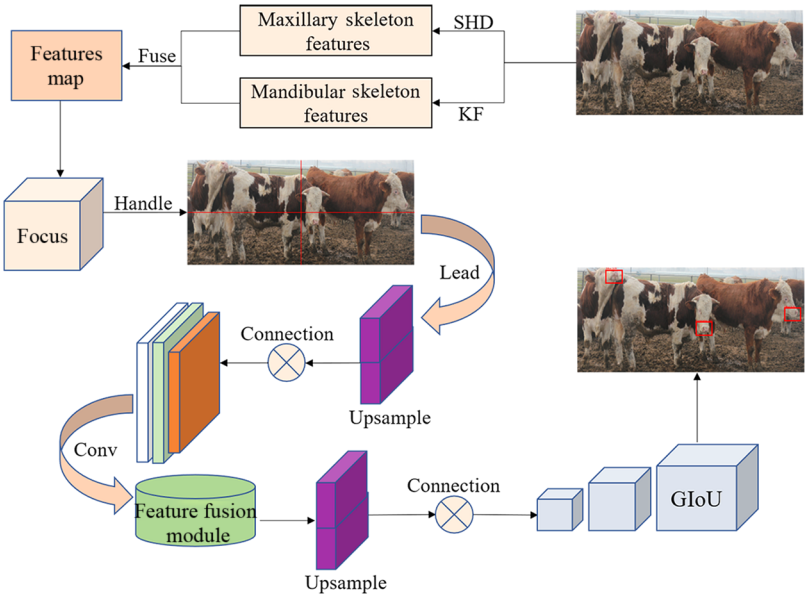

Fig. 1 System architecture diagram.

the jaw's motion features and skeleton information. The upper and lower jaw skeleton feature extraction captures distinct aspects of the cow's ruminant behavior. Precisely, the upper jaw skeleton feature reflects the morphological characteristics of the cow's mouth while the lower jaw skeleton feature reveals the movement characteristics of the cow's lower jaw. A more comprehensive and integrated representation of ruminant behavioral features is achieved by fusing these two features. As a target detection network, the YOLOv5s network can accept multichannel inputs. Therefore, the upper and lower jaw skeleton features can be effectively utilized as network inputs to achieve a comprehensive feature representation and analysis. This enhances the accuracy and effectiveness of our method for accurately identifying cattle grazing behavior.

3.2 UD-YOLOv5s Method Design

3.2.1 *Upper and lower jaw skeleton feature extraction*

Skeleton extraction is a valuable technique to detect essential points within an image and connect them sequentially, forming skeleton information. In the specific case of the cattle mouth region, the skeleton can be divided into two parts: the upper jaw and the lower jaw skeletons. The area of interest is first segmented into the upper and lower jaw regions to label the feature points in the original image. The feature points in the upper jaw are denoted as U1 to U6, corresponding to their coordinate positions starting from the left of the mouth area. Similarly, the feature points in the lower jaw are labeled as D1 to D5. Each pair of feature points between the upper and lower jaw is represented by a unit vector, such as U1→D1, connecting the respective points. However, due to the absence of a corresponding feature point for U6 in the lower jaw region, it is excluded from the skeleton. Figure 2 visually shows the labeled vital points in the cattle's mouth region, offering an overview of the identified feature points.

In this paper, we have employed the SHD detector, a highly accurate tool with a 90.90% success rate in identifying crucial points on the head, showcasing exceptional detection precision. To tackle the challenges posed by demanding conditions such as occlusion caused by cattle enclosures, complex environments, and overlapping bodies, we have innovatively utilized an 8-layer stacked hourglass network (8SHD). This network effectively transforms the original image into a feature map, creating a robust skeletal extraction model. The architecture of this model is represented in Fig. 3.

This paper presents an advanced algorithm designed to extract crucial features of the mandible skeleton in cattle, which is vital in complex motion patterns. The method involves a multistep process to ensure accuracy and reliability. First, the algorithm filters out low-confidence vital points, ensuring that only the most reliable maxillary structure critical points are considered.

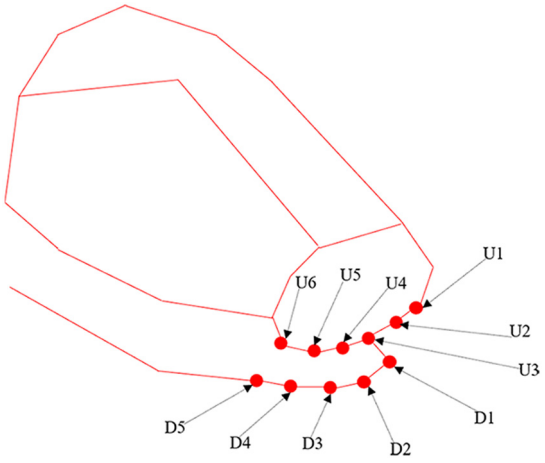

Fig. 2 Point marker map of the bull's mouth region.

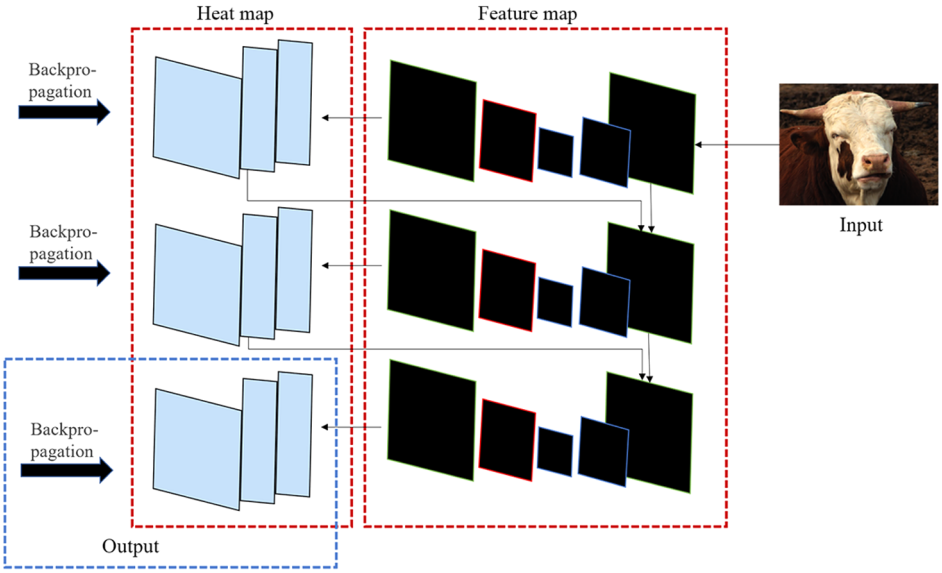

Fig. 3 SHD model diagram.

This selection is based on whether the output heat map exceeds a predefined threshold. An elliptical curve is employed to approximate the mandibular region, which aids in accurately capturing its shape. The algorithm then proceeds to identify vital points based on specific characteristics of the mandibular skeleton. These points serve as the foundation for further analysis. Next, a predictive algorithm is utilized to set key point features, refining the representation of essential points in the mandible. The algorithm considers the dynamic nature of the mandibular motion sequence, treating it as a dynamic system. During the opening and closing of the mandible, the left mouth edge point is selected as the original feature point, acting as a reference for the entire motion sequence. The curve formation is thoughtfully arranged by incorporating the mandibular skeleton feature and setting the adjacent bone distance to 0.1 cm per unit distance. In addition, the neighboring nodes of each unit period are considered adjacent skeleton feature points. By employing this sophisticated approach, the algorithm accurately models the motion pattern of the mandible.

When AP $= 1$ is recorded as the initial frame K, at this time, the jaw joint point B_k gets the anchor frame and sits marked as (x_1, y_1). Currently, the next frame joint point $B_{k|k+1}$ position can be predicted by the KF algorithm as

$$B_{k|k+1} = A(x_1, y_1), \tag{1}$$

where A denotes the state transfer matrix, and the site-tag of the next frame B_{k+1} is predicted to be (x_2, y_2). The covariance matrix is calculated according to B_k and B_{k+1} prediction variable densities as

$$P_{k|k+1} = AP_{k+1|k+2}A^T + Q, \tag{2}$$

where $P_{k|k+1}$ is denoted as the covariance matrix of B_k, $P_{k+1|k+2}$ is represented as the covariance matrix of $B_{k|k+1}$, Q is the process noise, and T is the time stamp. The position of the detected node after processing is determined with the predicted position, and the updated state is calculated by Kalman gain. The coordinates of the matching node are replaced with the predicted coordinates. The gain is calculated as

$$K_k = P_{k|k-1}H^T(HP_{k|k-1}H^T + V)^{-1}. \tag{3}$$

The replacement state after gain is calculated as

$$W_{k|k} = W_{k|k-1} + K_k(X_k - Hx_{k|k-1}). \tag{4}$$

The error covariance is shown as

$$P_{k|k} = (1 - K_kH)P_{k-1|k-1}, \tag{5}$$

where K_k denotes the Kalman gain of frame k; W_k denotes the updated lower jaw position; P_k denotes the covariance matrix. We denote the obtained initial key point of the upper jaw as p1, the next predicted node of the frame as p2, and the adjacent algorithm-predicted nodes as p3 and p4. The initial key point of the lower jaw as p5, the next predicted vital point of the frame by the algorithm as p6, and the adjacent nodes as p7 and p8. The skeleton critical point representation is shown in Fig. 4.

Information setting based on the jaw key point representation. Key point parameters are shown in Table 1.

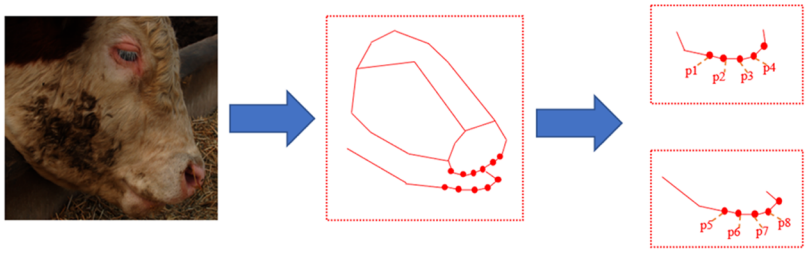

Fig. 4 Cattle skeleton key point representation diagram: (a) original image, (b) key point extraction, and (c) key point of the jaw.

Table 1 Key point information based on algorithm settings.

Key point representation	Key points description
P1	Initial trace box key points
P2	Next frame forecast point
P3	Next frame forecast point
P4	Next frame forecast point
P5	Initial trace box key points
P6	Next frame forecast point
P7	Next frame forecast point
P8	Next frame forecast point

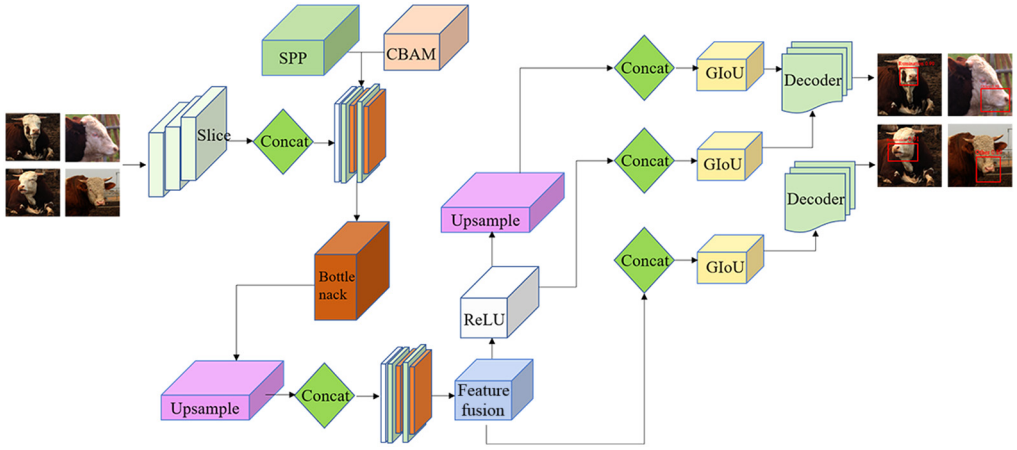

Fig. 5 UD-YOLOv5s network structure diagram.

3.2.2 *Design of UD-YOLOv5s network*

The YOLOv5s object detection algorithm is renowned for its exceptional detection accuracy and lightweight architecture. In this research, we introduce a groundbreaking approach called the UD-YOLOv5s network, specifically designed for detecting the mouth area of cattle and recognizing their rumination behavior. To achieve this goal, our methodology begins by resizing the input image to 640 px × 640 px. Subsequently, we employ a slice operation to divide and connect the image, generating a processed image for further analysis. We incorporate spatial pyramid pooling and the convolutional block attention module (CBAM) to enhance the receptive field and feature expression. These components capture essential information from various spatial scales, optimizing the network's performance. The data are then passed to the convolutional layer, where a bottleneck mechanism is employed for efficient feature extraction, thereby reducing the computational burden. The subsequent step involves feeding the up-sampling connected convolutional layer into the feature fusion module. This integration is executed bottom-up, employing the ReLU activation function, up-sampling connection, and GIoU loss function for effective optimization. Finally, the output of the feature map includes target bounding boxes, class probabilities, and confidence scores. A decoder completes this process, facilitating the precise identification and localization of cattle mouth regions. The architecture of the UD-YOLOv5s network is visually shown in Fig. 5, providing an overview of our proposed model's intricate yet powerful design.

In the context of bull's mouth image analysis, addressing potential node mismatches that may occur when capturing images using a camera is essential. To ensure the accuracy and reliability of the skeleton features, we employ the object key point similarity (OKS) as a metric for evaluation, as shown in Eq. (6). OKS is a crucial measure to assess the similarity between the detected critical points in the image and the ground truth positions of these key points. Using OKS as an evaluation method, we can quantitatively gauge the alignment between the skeleton features extracted from the image and the expected positions of these features. The precise definition and calculation of OKS are shown in Eq. (6), allowing us to effectively evaluate and validate the correctness of the identified skeleton features compared to the ground truth. By incorporating the OKS evaluation method, we can enhance the robustness of our bull's mouth image analysis, ensuring that the skeleton features accurately represent the underlying anatomical structures even in scenarios where node mismatch may occur during image capture

$$O_{ks} = \frac{\sum_i e^{-\frac{di}{2Sab}}}{\sum b}, \tag{6}$$

where O_{ks} denotes the similarity of the key points of the maxillary skeleton, a, b represents the visibility of the critical points, S the pixels of the individual feature points of the skeleton, d_i the Euclidean distance between the actual coordinates of the critical points and the predicted coordinates, and i the normalization factor of the key points. After inputting the regurgitated test set into the model, the average precision (AP) is calculated by calculating the OKS ratio to the

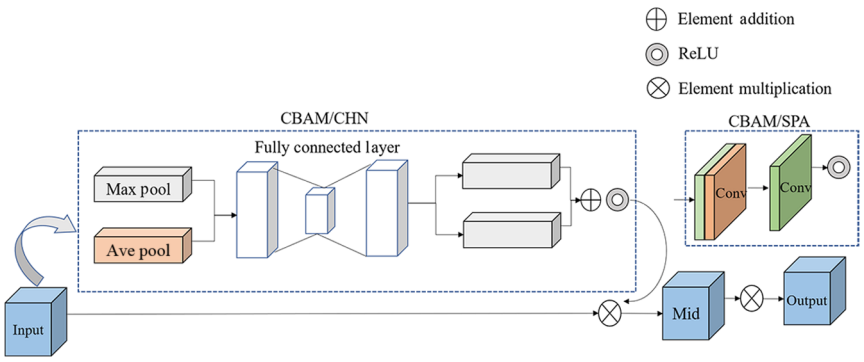

Fig. 6 CBAM attention mechanism structure.

skeleton individual node threshold, and determining whether the predicted target detection node coordinates match the actual coordinate points, if AP = 0, the current image detection error is determined and the next image is input; AP = 1, the skeleton node anchor frame is generated by the adaptive anchor frame mechanism, and the results are output to the backbone layer.

The backbone network layer and neck network layer are our model's core components and have been optimized using a lightweight approach. To enhance detection accuracy, we have incorporated the CBAM into the backbone layer for the skeleton nodes. CBAM comprises two main modules: the channel attention (CHN) module and the spatial attention (SPA) module. By leveraging CHN, the model can focus on relevant features within each channel, effectively improving the overall representation. Subsequently, the feature map is passed through the SPA, enabling the model to extract crucial spatial information, further enhancing feature extraction. For a visual representation of the CBAM attention mechanism, refer to Fig. 6.

In CHN module, the original feature map X, which is the product of height H, width W, and number of channels N, is pooled to obtain the channel map, and then processed by multilayer perceptron (MLP) to obtain the feature weights, combined with the ReLU activation function to obtain the channel weight coefficients M_n. The product of M_n and M is used as the scaled channel feature map Y, and M_n is calculated as

$$M_n = \partial(\mathrm{MLP}(\mathrm{Pool}(X))), \tag{7}$$

where ∂ represents the ReLU activation function, MLP the multilayer perceptron, and Pool the adaptive pooling operation. In the SPA module, the new feature map Y is pooled to obtain the channel map for stitching, and the spatial weight coefficient M_s is obtained after $3 * 3$ layer convolution and ReLU activation function, and the product of M_s and the feature map Y is used as the output feature map Z. The formula of H is as in Eq. (8)

$$M_s = \theta^{3*3}(\mathrm{Pool}(Y)), \tag{8}$$

where θ^{3*3} means the $3 * 3$-layer convolution and Y the channel feature map. The filtered feature map Z is input to the Neck layer, as shown in Fig. 7.

The neck layer employs a bottom-up path aggregation feature pyramid to transfer feature information efficiently. This pyramid facilitates the integration of feature maps corresponding to key points' information. The process involves convolving the feature map of each skeleton key point with a step size of 1, which effectively fuses the features. Up-sampling techniques are utilized to accomplish the feature fusion, and the convolution module consists of a series of

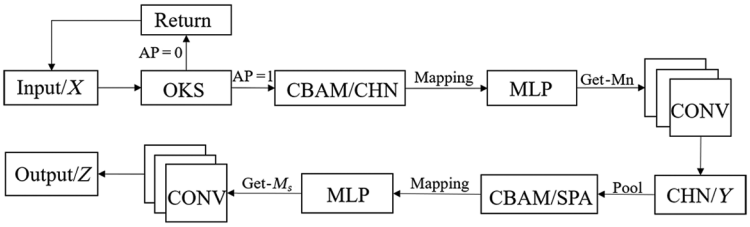

Fig. 7 UD-YOLOv5s feature extraction map.

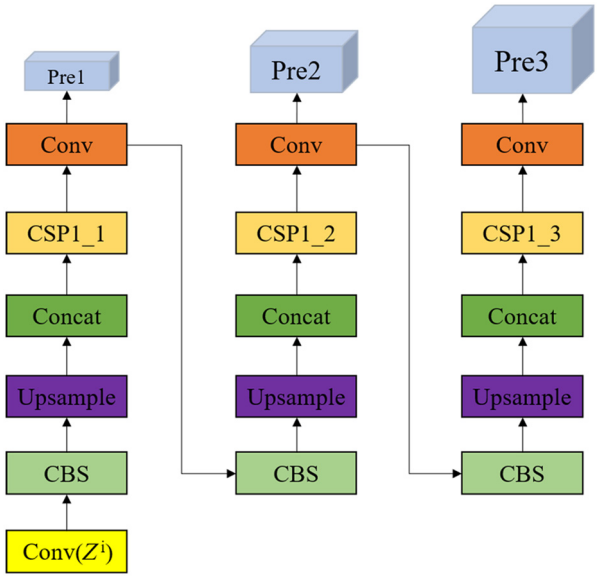

Fig. 8 UD-YOLOv5s path aggregation neck network architecture diagram.

x residual structures resource units. This integration process enhances the overall representation and prepares the feature information for the subsequent detector. Ultimately, the prepared feature information is fed into the detector, where regurgitated behavior recognition occurs, as illustrated in Eq. (9)

$$V = \begin{cases} \text{Pool}(Z^i)^{1*1} \rightarrow \text{Upsample} \\ \text{Prediction}(\text{Mod}^x) \end{cases}, \tag{9}$$

where V represents the feature information, Upsample the up-sampling operation, Mod^x the convolution module integrated by x residual structures, and Prediction the detector prediction operation. The neck layer architecture and output are shown in Fig. 8.

The output tensor is partitioned into multiple small tensors, each corresponding to a network cell and its surrounding prediction frame. Within each small tensor, computations are performed to determine the target confidence, category probability, and bounding box coordinates. The final prediction takes the form of a three-dimensional tensor, where the first dimension represents the identification of each grid cell, the second dimension means the identification of each predefined frame, and the third dimension contains information related to target confidence, category probability, and bounding box coordinates. The upper and lower jaw skeleton feature extraction method is employed to obtain multiple anchor frames in the mouth region. This technique can derive anchor frames based on the skeleton features of the upper and lower jaws. Specifically, the center point of the upper jaw anchor frame serves as the base point, and it is connected to the center point of the lower jaw anchor frame. The Euclidean distance between these two anchor points is then calculated as

$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}, \tag{10}$$

where x_1, x_2, y_1, and y_2 denote the horizontal coordinates of the center point of the upper and lower jaw anchor frames and the vertical coordinates of the center point of the upper and lower jaw anchor frames, respectively. Take the minimum value of $\text{Min}(d^K)$, at this time k frame is recorded as a chewing determination, regurgitation is consistent with chewing periodicity, so the determination of the number of regurgitation α as shown in Eq. (11)

$$\alpha = l/(\text{Min}^k - \text{Max}^{k+1}), \tag{11}$$

where Min^k denotes the minima on the mastication curve in frame k, Max^{k+1} the maxima in the adjacent next frame, and l the mastication curve length. The final recognition prediction output is completed in the prediction layer.

3.3 Optimization of UD-YOLOv5s

Recognizing ruminant behavior poses a significant challenge due to an inherent dataset imbalance, with a limited number of ruminant samples compared to an abundance of non-ruminant instances. Traditional YOLOv5s rely on the IoU loss function, which needs to be more accurate in capturing the overlap between the target and authentic frames. This paper tackles this issue and aims to enhance model learning by introducing the GIoU loss function, offering a more precise measurement of frame overlap. We note that we retained all data in our work to maintain dataset integrity.

3.3.1 *Loss weighting*

GIoU contains both localization error and classification error, so we increase the corresponding loss weights, and the localization error weight (W_{i_loc}) and classification error (W_{i_cls}) weights are calculated explicitly as

$$W_{i\_\mathrm{loc}} = N/(N_i \times 2), \tag{12}$$

$$W_{i\_\mathrm{cls}} = 1 - W_{i\_\mathrm{loc}}, \tag{13}$$

where N denotes the category and N_i the number of positive samples of the i sample. We see that the fewer the category samples, the larger the corresponding loss weight.

3.3.2 *Positive and negative sample sampling*

In the context of a relatively small sample size of regurgitation behavior, the imbalance between positive and negative samples can adversely impact model training, leading to suboptimal results. We propose a positive and negative sample improvement strategy based on UD-YOLOv5s to address this issue. This strategy comprises three essential steps: redefining positive and negative samples, implementing a problematic sample mining approach, and enhancing the sample balancing strategy using the gradient harmonized (GHM) single stage detector loss for the GIoU loss function.

Step 1: Redefining positive and negative samples.

In the first step, we redefine positive samples as prediction boxes containing cattle while considering all other instances as negative samples. By focusing the model's attention on cattle-containing boxes, we aim to effectively improve its ability to detect regurgitation behavior.

Step 2: Problematic sample mining strategy.

We employ a problematic sample mining strategy to enhance the model's learning rate. During data collection, we introduce some challenging-to-identify samples to the negative samples and then incorporate these additional instances into the training set. This approach helps the model better handle challenging scenarios, leading to improved performance.

Step 3: Improving the sample balancing strategy.

In the final step, we augment the sample balancing strategy by incorporating the GHM. This strategy ensures a more balanced calculation of the GIoU loss function by dividing gradient values into intervals using quantiles as division points, forming bins in the gradient histogram. Subsequently, the value of each sample is calculated using the defined methodology as shown in Eq. (14). By adopting these three steps in our positive and negative sample improvement strategy, we aim to overcome the challenges posed by the limited regurgitation behavior dataset and achieve more accurate and robust results in cattle regurgitation behavior detection using UD-YOLOv5s

$$\mathrm{GIoU}_{\mathrm{loss}}(p_i, t_i) = \sum_{i=1}^{C} w_i |p_i - t_i|, \quad w_i = \frac{1}{\sqrt{\sum_{j=1}^{k} g_j}}, \tag{14}$$

where p_i and t_i denote the predicted and actual scores, respectively, C the total number of samples, and g_j the number of samples within the j bin of the gradient histogram.

3.3.3 *Dynamic weight adjustment*

We adjust the weight of the loss function for that class according to the magnitude of the calculated loss value, and set each sample weight in UD-YOLOv5s as

$$w_i = \begin{cases} \alpha(1-q_i)^{\beta} \\ (1-\alpha)q_i^{\beta} \end{cases}, \tag{15}$$

where q_i denotes the prediction result of sample i, α and β are hyperparameters to implement the imbalance problem of GIoU loss function. The procedure of the UD-YOLOv5s learning algorithm is shown in Table 2.

4 Methods and Materials

4.1 Data Acquisition

The experiment was conducted at a ranch in Yuan Yang, Henan Province, China, involving 120 mature Simmental cattle. To ensure comprehensive coverage of cattle activities, a camera was placed at a 45 deg angle above the ranch. Ranch caretakers meticulously screened the video data, and instances of cattle ruminating within 3 h of feeding at 5 AM and 5 PM were selected. Cattle were observed to be resting quietly during other periods. Subsequently, a frame-by-frame analysis was performed to process the video frames into image datasets. These resulting data files were accurately labeled based on their contents, with a resolution of 640 px $*$ 480 px. Both training and testing utilized the captured cattle behaviors. In addition, different degrees of occlusion were

Table 2 UD-YOLOv5s algorithm.

UD-YOLOv5s algorithm
INPUT: X // Original feature map
OUTPUT: //z
Function UD-YOLOv5s () {
FOR ($i = 1$; $i <= n$; $i++$)
{
$O_{ks} = \sum_i e^{-\frac{di}{2S_{ab}}} / \sum b$ // OKS metrics assessment
IF (AP = 1) {
$Mn = \partial(\text{MLP}(\text{Pool}(X)))$ // Channel feature weights
$Ms = \theta^{3*3}(\text{Pool}(Y))$ // Spatial feature weights
$V = \begin{cases} \text{Pool}(Z^i)^{1*1} \rightarrow \text{Upsample} \\ \text{Prediction}(\text{Mod}^x) \end{cases}$ //Extracting feature information V
$B_{k\|k+1} = A(x1, y1)$ // Next frame position prediction
$P_{k\|k+1} = \text{AP}_{k\|k+1}A^T + Q$ // Covariance matrix
$d = \sqrt{(x1-x2)^2 + (y1-y2)^2}$ // Mastication curve determination
$\alpha = I/(\text{Min}^k - \text{Max}^{k+1})$ // Calculating the number of regurgitations
ELSE
CONTINUE
ENDIF
}
ENDFOR
}
}

Table 3 Definition of cattle ruminant labeling.

Behavior	Define
Rumination	Chewing behavior of cattle while standing, lying, sitting, etc., after the head leaves the trough.
Other	Cattle standing, sitting, eating and drinking, and other behaviors.

incorporated into the data collection process to reflect real-life farm conditions. The definition of cattle rumination is shown in Table 3, ensuring transparency and replicability of the work.[44]

4.2 Data Pre-Processing

In this work, the images utilized were impacted by low quality, mainly due to overlapping shapes caused by obstructions from the farm cattle pen, cattle movement, and poor adaptability to varying lighting conditions. To ensure data integrity, these low-quality images were excluded from the dataset. To enhance the generalization ability of the recognition network, data augmentation techniques were employed. Rotation and the mosaic data augmentation technique effectively augment the dataset. In addition, classic data augmentation techniques, such as image flipping, scaling, cropping, translation, noise addition, and contrast modification, were applied to diversify the dataset further. For a visual representation of the augmented dataset, refer to Fig. 9. These augmentation approaches collectively improve the model's ability to recognize cattle behaviors amidst challenging environmental conditions and variations in image quality.

Mosaic enhancement is a widely adopted image processing technique in object detection. In the YOLOv5s network, the cut and mix data augmentation algorithm is employed, effectively expanding the dataset and enhancing its diversity.[45,46] Similarly, in UD-YOLOv5s, the stitcher method is incorporated as a data augmentation approach to tackle the issue of low accuracy in object detection caused by poor image quality resulting from various environmental factors during image capture.[47] Figure 10 visually shows the process diagram for the mosaic enhancement technique employed in UD-YOLOv5s. By leveraging the mosaic enhancement and stitcher methods, the model's ability to detect and recognize objects, even in challenging conditions with compromised image quality, is substantially improved. These techniques are pivotal in optimizing the performance of UD-YOLOv5s for object detection tasks.

The mosaic-enhanced image of the data in this paper is shown in Fig. 11.

In this context, consider Figs. 11(a1)–11(c1) as the original images and Figs. 11(a2)–11(c2) as the processed images. For data annotation, we utilized the labeling tool, and the resulting dataset is named the "cattle dataset." This dataset comprises two distinct categories: "ruminate" and "other," it encompasses a total of 1500 processed datasets. The data annotation details can be observed in Fig. 12. The use of the labeling tool ensures accurate and consistent annotation.

Fig. 9 Cattle expansion data chart: (a1)–(c1) original image, (a2) image denoising, (b2) image cropping, and (c2) image scaling.

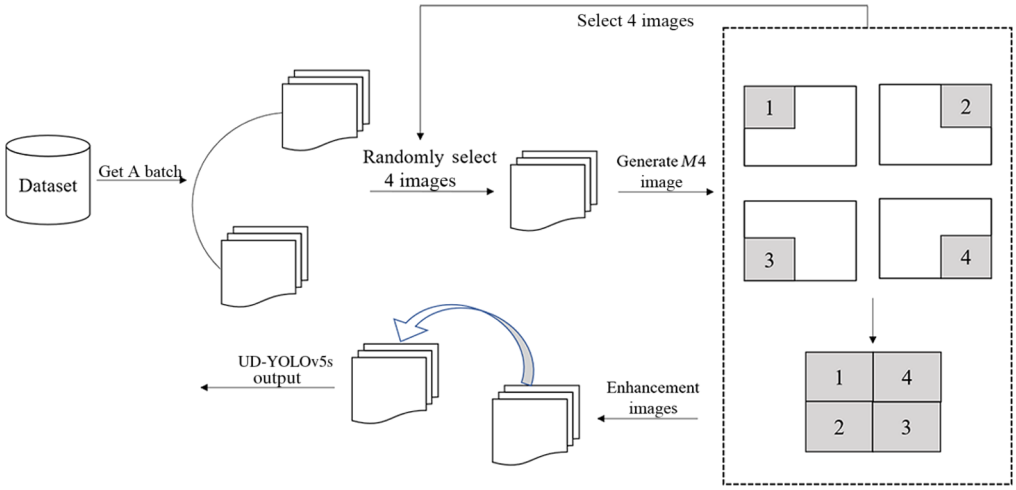

Fig. 10 Mosaic data enhancement flowchart for UD-YOLOv5s.

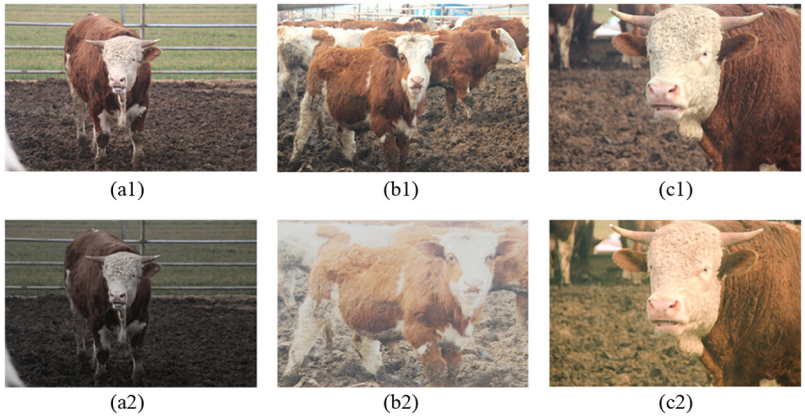

Fig. 11 Mosaic image enhancement map: (a1)–(c1) original image, (a2) brightness adjustment, (b2) contrast adjustment, and (c2) saturation adjustment.

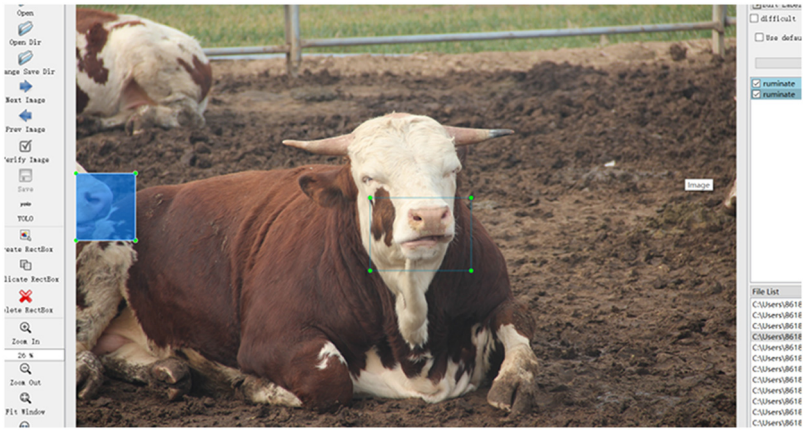

Fig. 12 Data annotation chart.

At the same time, the division of the dataset into "ruminate" and "other" categories facilitates targeted analysis of cattle behavior. This comprehensive dataset lays the foundation for robust training and evaluation of the model, enabling effective detection and recognition of cattle activities.

4.3 Experimental Environment and Parameter Settings

The experiments were conducted on a LINUX operating system, utilizing a robust hardware setup consisting of 16 GB of RAM, an NVIDIA GEFORCE RTX3070 graphics card for GPU acceleration, and a Core i7 CPU, along with an Intel(R) Core (TM) i7-10750H CPU@2.60 GHz 2.59 GHz processor for network training configuration. The model was built using the PyTorch deep learning framework, known for its efficiency and flexibility. Careful attention was given to the hyperparameter settings to ensure optimal performance during the training phase, as outlined in Table 4. These hyperparameter configurations play a crucial role in shaping the model's learning process, enabling it to capture intricate patterns and features in the cattle dataset effectively. The robust hardware setup and well-tuned hyperparameters contribute significantly to the model's accuracy and efficiency, enabling comprehensive training and evaluation of the object detection model for cattle behavior recognition.

To thoroughly evaluate the efficacy of the YOLOv5s network in detecting the upper and lower jaw regions of cattle during rumination behavior, we conducted an extensive rumination recognition detection experiment using the cattle dataset introduced in this paper. To ensure a fair and comprehensive comparison, we carefully selected MEAN-SHIFT, MASK-RCNN, and YOLOv3 as the control group networks, considering their similarities, versatility, and advancements in the field.

MEAN-SHIFT, a clustering algorithm, segmented images by analyzing pixel color distribution. We extracted mouth features from video sequences and performed clustering operations using MEAN-SHIFT to recognize cattle rumination behavior occurring in the mouth area.

MASK-RCNN, an enhanced algorithm capable of simultaneous object detection and instance segmentation, played a vital role in the experiment. We labeled the cattle rumination

Table 4 Hyperparameter setting.

Parameters	Numerical value
Cuda	10.0.1
CuDNN	10.0.1
Pytorch	1.2.0
Initial learning rate	0.01
Termination rate	0.2
Number of prediction frames/pc	3
Learning rate decay strategy	Cosine annealing
Data enhancement methods	Mosaic enhancement
Number per input	10
Momentum parameters	0.9
Input pixels	608×608
Final decay rate	4.9×10^{-4}
Batch	32
Number of training sessions	100
Momentum factor	0.932
Number of categories	80

Table 5 Neural network parameter setting.

Parameters	Value
Anchor boxes	3
Batch size	1000
Output layer activation function	ReLU
Loss function	GIoU
Feature extractor	CSPDarkent53
Learning rate scheduler	StepLR scheduler
Hidden layer activation function	Mish
Epoch	100

dataset and utilized pre-trained models to detect and locate the mouth area of the cattle. The region information was then fed into the MASK-RCNN network to perform mask segmentation, generating a segmentation mask for the cattle's mouth. Subsequently, we extracted the feature vector and performed classification to identify cattle rumination behavior.

The YOLOv3 algorithm involved video processing software to extract images containing cattle rumination behavior and conducted object detection to obtain the position information of the cattle's mouth. We pulled relevant feature information, such as motion trajectory and color, and employed statistical analysis to determine the presence of rumination behavior.

To ensure a fair comparison, we maintained consistent hyperparameters across all networks and presented identical parameter configurations for the four networks, as shown in Table 5. This approach allowed us to draw comprehensive conclusions and thoroughly evaluate the performance of UD-YOLOv5s in cattle rumination behavior recognition tasks from multiple perspectives. By conducting a rigorous comparative analysis, we aimed to provide robust insights into the strengths and weaknesses of each algorithm, ultimately highlighting the effectiveness of UD-YOLOv5s in this domain.

Since MEAN-SHIFT is a density-based unsupervised clustering algorithm, MASK-RCNN, YOLOv3, and UD-YOLOv5s are based on CNN architecture, the parameters of the four reconstructed networks are shown in Table 6.

5 Experiment

5.1 UD-YOLOv5s Ablation Experiment

We divided the dataset into "rumination" and "other," represented by labels 0 and 1, respectively. For model training and validation, we utilized a 9:1 partition ratio. The training phase comprised 100 cycles to optimize model performance. To thoroughly evaluate the effectiveness of the jaw skeleton feature extraction method, we conducted ablation experiments under identical experimental conditions. In this paper, using the cattle dataset, we compare and validate the performance of four approaches: MEAN-SHIFT, MASK-RCNN, YOLOv3, and UD-YOLOv5s. The critical evaluation metrics for the models encompass precision (Pre), recall (Rec), frames per second (FPS), mean AP (mAP), model memory usage (MMU), and $F1$ score. Pre denotes the ratio of samples predicted as positive by the genuinely positive model. In the context of cattle rumination recognition, it measures the accuracy of images correctly classified as rumination or non-rumination. Rec signifies the proportion of positive samples the model accurately predicts, indicating the number of samples correctly classified as positive by the classifier divided by the actual number of positive samples. It provides insight into how effectively rumination images are identified. The calculation method for the recall is illustrated in Eqs. (16) and (17). In this comprehensive evaluation, these metrics serve as key performance indicators, shedding light on the models' abilities to detect cattle rumination behavior accurately. The higher the precision, recall, and $F1$ score, the more reliable and effective the model recognizes cattle rumination patterns. In addition, the FPS, mAP, and MMU metrics contribute valuable insights into each model's

Table 6 Reconfigured network parameters.

Network	Parameters	Value
MEAN-SHIFT	Window size	10
	Kernel function	Gaussian function
	Number of iterations	100
	Convergence threshold	0.001
YOLOv3	Input size	416 × 416
	Learning rate	0.001
	Regularization factor	0.0005
	Batch size	64
	Network structure	Darknet-53
	Activation function	Leaky ReLU
MASK-RCNN	Feature extraction network	101 convolutional layers
	Area proposal network	3 candidate boxes
	Classification network	2
	RoI pooling layer	14 × 14
UD-YOLOv5s	Input size	608 × 608
	Learning rate	0.01
	Regularization factor	0.0005
	Batch size	16
	Network structure	CSPNet
	Activation function	ReLU

CSPNet, cross stage partial network.

computational efficiency and memory utilization. These evaluations are crucial in determining the most suitable approach for cattle rumination recognition tasks

$$\mathrm{Pre} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}, \tag{16}$$

$$\mathrm{Rec} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}, \tag{17}$$

where true positive (TP) represents the number of correctly identified positive cases, false positive (FP) indicates the number of FP claims, and false negative (FN) represents the number of FN claims. After conducting a thorough experimental analysis, UD-YOLOv5s demonstrates outstanding performance, achieving a remarkable 98.25% accuracy in bovine regurgitation recognition. This accuracy is 2.08% higher than that of YOLOv3, 7.22% higher than MASK-RCNN, and an impressive 10.18% higher than the MEAN-SHIFT algorithm. Regarding recall, UD-YOLOv5s showcases remarkable results, reaching a 97.75% recall rate. This is 4.07% higher than YOLOv3, 0.63% higher than MASK-RCNN, and 11.1% higher than the MEAN-SHIFT algorithm. The significant accuracy and recall improvements achieved by UD-YOLOv5s demonstrate its effectiveness and potential for real-world cattle rumination behavior detection tasks. The $F1$ score is the weighted average of accuracy and recall, which can combine the model's accurate prediction and completeness prediction and tells us that the model can accurately predict regurgitated pictures while avoiding missed detection as much as possible. The mAP indicates the degree of match between all boxes predicted by the model and the authentic boxes under

different confidence thresholds. It is used to judge the overall performance of all models on regurgitated behavior recognition performance. It is calculated as shown in Eqs. (18) and (19)

$$F1 = \frac{2 \times \text{Pre} \times \text{Rec}}{\text{Pre} + \text{Rec}}, \tag{18}$$

$$\text{mAP} = \frac{(\text{AP}_1 + \text{AP}_2 + \cdots + \text{AP}_n)}{n}, \tag{19}$$

where AP denotes the average precision and n indicates the total number of categories. The experimental results showed that UD-YOLOv5s reached 97.59%, 2.37% higher than YOLOv3, 3.27% higher than MASK-RCNN, and 11.92% higher than MEAN-SHIFT. Since the MEAN-SHIFT algorithm needs to dichotomize the results in the bovine rumination task, the mAP metric does not apply to this algorithm, and we did a mAP evaluation of the other three algorithms, which showed that UD-YOLOv5s reached 93.43%, 6.19% higher than YOLOv3, and 13.54% higher than MASK-RCNN. FPS refers to the number of frames a computer can process per second when analyzing video data. It serves as a crucial metric to gauge the real-time performance of a model in cattle ruminant recognition tasks, providing insights into the model network's low latency and overall recognition improvement. A higher FPS value indicates faster model processing, resulting in shorter response times and more efficient real-time performance.

On the other hand, MMU quantifies the memory size occupied by a trained deep-learning model. In the context of cattle ruminant recognition tasks, MMU serves as a measure of the model's complexity. A smaller MMU implies a more compact model footprint, making it feasible to deploy the model on smaller devices for inference. This can help avoid overfitting and enhance the model's generalization performance as it becomes more adaptable to various hardware configurations. The calculation of MMU is typically carried out following the equations provided in Eqs. (20) and (21)

$$\text{FPS} = \frac{1}{t_{\text{image}}}, \tag{20}$$

$$\text{MMU} = S \times \text{Bit}, \tag{21}$$

where t_{image} denotes the time required to process a batch of images, Bit indicates the number of bytes per parameter data type, and S denotes the number of model parameters. The experiments show that UD-YOLOv5s has a higher FPS with the same training time and a smaller model memory footprint with the same parameters and weights loaded. The results of the experiment are shown in Table 7.

5.2 Evaluation of the Generalization Effect of UD-YOLOv5s

We divided our self-built dataset into two main parts: the training dataset and the test dataset. To assess the generalization effect of the UD-YOLOv5s model, we conducted a thorough evaluation by comparing the accuracy and loss rate between these two sets. The dataset was further categorized into two groups: the regurgitated dataset and the other datasets which were used in the experiment. For the crossover subjects, we ensured that the training and test sets were appropriately balanced, consisting of 16,523 and 7152 samples, respectively. Similarly, for the crossover view, we maintained a balanced distribution of 14,236 samples in the training set and 6482 samples in the test set. Figure 13 illustrates the insightful results derived from our evaluation. The

Table 7 The results of the experiment.

Model	Pre (%)	Rec (%)	AP (%)	FPS (f · s^{-1})	mAP (%)	MMU (MB)	$F1$ (%)
MEAN-SHIFT	88.07	86.65	85.67	10.21	—	235.67	81.19
MASK-RCNN	91.03	97.12	94.32	12.54	79.89	115.45	91.53
YOLOv3	96.17	93.68	95.22	13.35	87.24	108.74	92.17
UD-YOLOv5s	98.25	97.75	97.59	53.96	93.43	36.73	94.09

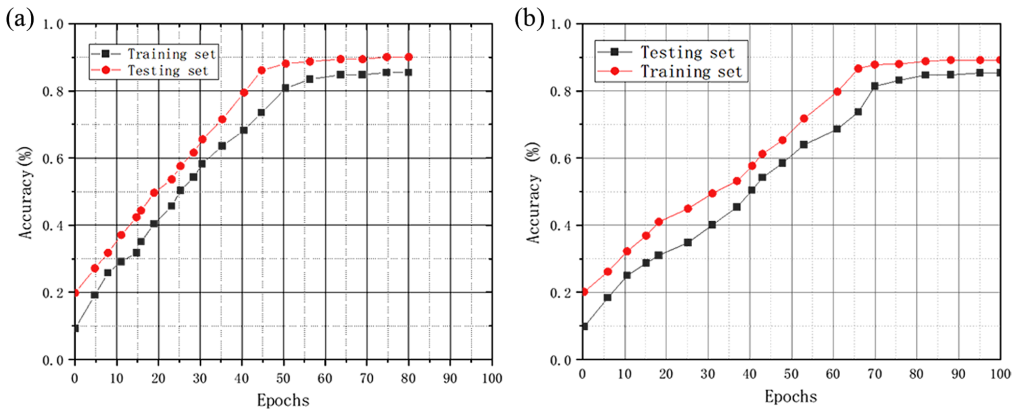

Fig. 13 Comparison of the accuracy of the training set and the test set: (a) cross subject (CS) and (b) cross view (CV).

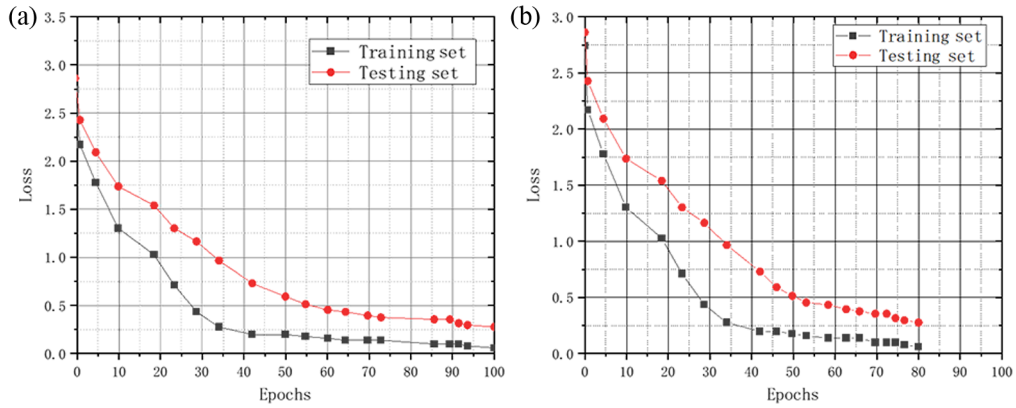

Fig. 14 Comparison of the loss values of the training set and the test set: (a) CS and (b) CV.

model's accuracy exhibits a consistent upward trend as the number of training samples increases, reflecting the positive impact of additional training data on enhancing the model's performance.

The loss rates associated with the training and test sets in both cross-subject and cross-view iterative training are shown in Fig. 14. The figure demonstrates a notable pattern: as the number of training samples increases, the model's loss rate consistently decreases. This observation highlights the beneficial effect of utilizing more training data, as it improves the model's ability to minimize errors and optimize its predictions. This reduction in loss rate reflects the model's enhanced learning and generalization capabilities, making it more adept at handling different scenarios and achieving better overall performance.

5.3 Performance Comparison

To showcase the outstanding performance of the UD-YOLOv5s model in rumination recognition, we conducted a comprehensive comparative analysis involving three state-of-the-art models: MEAN-SHIFT, MASK-RCNN, and YOLOv3, all of which have demonstrated exceptional capabilities in rumination recognition research. MEAN-SHIFT focuses on tracking the mouth area and extracting mouth features to detect rumination patterns. MASK-RCNN, on the other hand, harnesses the power of the residual neural network to extract mouth area features and trains the algorithm, for instance, segmentation, enabling precise rumination recognition. As for YOLOv3, it processes the image using the backbone network, effectively fuses the features, and produces results through the prediction layer for rumination recognition. Since rumination recognition, which relies on the extraction of mandibular and maxillary skeleton features, is a multi-classification task, we employed the GIoU loss function for model training. To assess the superior convergence efficiency of the UD-YOLOv5s model, we meticulously compared the

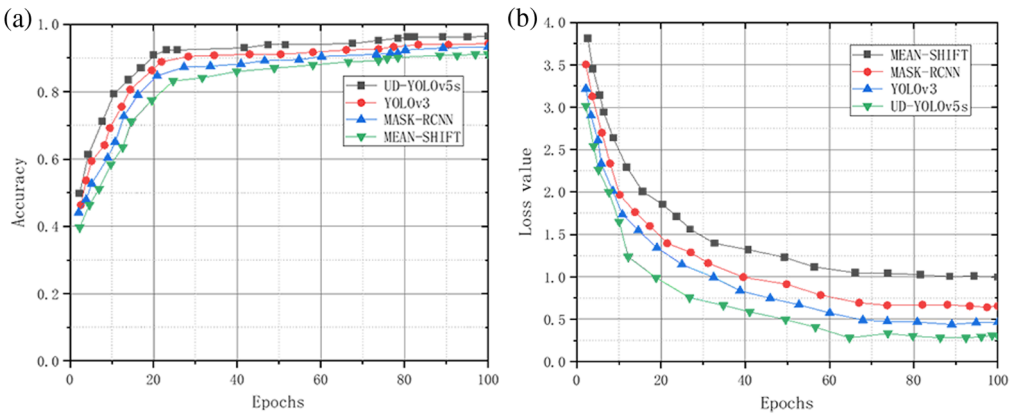

Fig. 15 Performance comparison of UD-YOLOv5s and other algorithms: (a) accuracy and (b) loss value.

accuracy and loss values of MEAN-SHIFT, MASK-RCNN, and YOLOv3 at varying training cycle durations, as visually shown in Fig. 15. The results were resoundingly in favor of UD-YOLOv5s, exhibiting a remarkably faster convergence speed while maintaining the same level of recognition accuracy as the other models. When compared at the same convergence speed, UD-YOLOv5s achieved significantly higher recognition accuracy. Consequently, the UD-YOLOv5s model demonstrates unparalleled performance in rumination recognition accuracy when contrasted with the other models subjected to scrutiny.

This work is based on the YOLOv5s, and the results of the experimental comparison with YOLOv5s are as follows in Table 8.

Based on the accuracy, recall, mAP, and $F1$ metrics, we conducted statistical tests on a self-constructed ruminant dataset using the following approach:

Data preparation: the regurgitated and non-regurgitated datasets used in the experiment were selected as samples for the statistical test.

Model evaluation: we compared and evaluated the different networks' accuracy, recall, mAP, and $F1$ scores.

Testing method: considering the nature of the cattle ruminant dataset and the fulfillment of statistical assumptions, we chose the t-test as the statistical testing method.

Hypothesis determination: we established the null hypothesis ($H0$) and alternative hypothesis ($H1$). The null hypothesis states that there is no significant difference between the indicators of the two models. In contrast, the alternative view suggests a significant difference exists between the indicators of the two models.

Statistical calculations and significance tests: we performed t-tests to calculate and assess the significance of the indicator differences. Based on the chosen method and the calculated results, the obtained p-values were used to determine whether the differences were statistically significant.

The experimental results demonstrated that the calculated p-value was less than the chosen significance level ($\alpha = 0.05$), allowing us to reject the null hypothesis and conclude that the observed differences are statistically significant.

We analyzed the impact of UD-YOLOv5s on the interpretation of actual farm cattle regurgitation videos, as shown in Fig. 16. In the context of recognizing cattle regurgitation behavior in

Table 8 YOLOv5s comparison experiment results.

	Evaluation index (%)			
Model	Precision	Recall	mAP	$F1$ score
YOLOv5s	96.31	95.98	91.87	92.66
UD-YOLOv5s	97.43	97.21	92.74	93.87

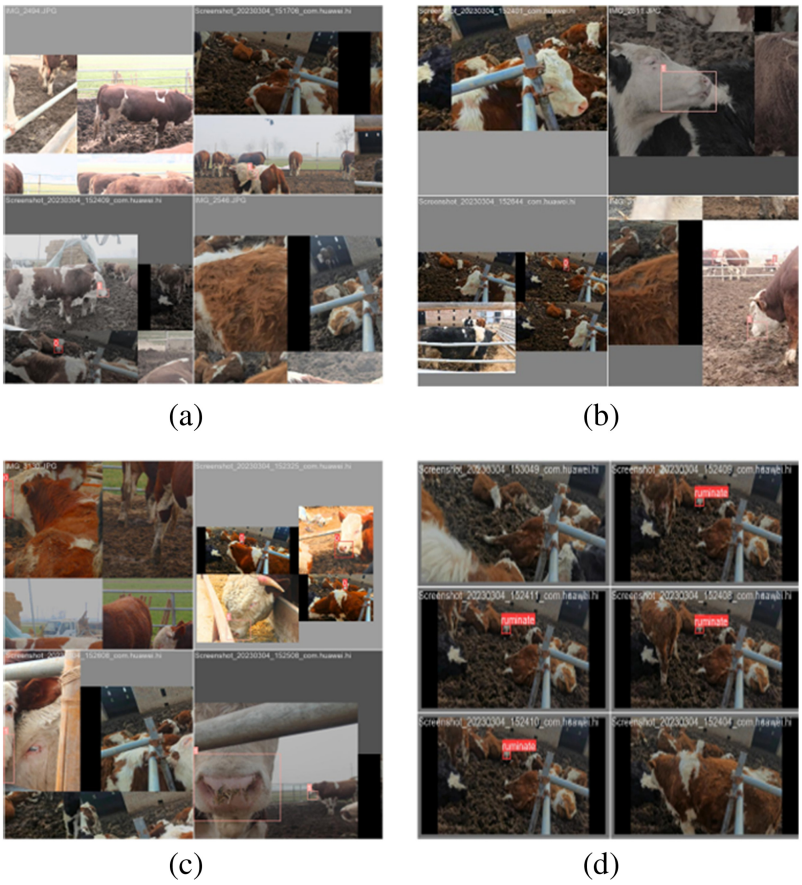

Fig. 16 Graph of experimental results of UD-YOLOv5s: (a), (b) other behavior and (c), (d) ruminant behavior.

various scenarios, it is evident that the recognition method effectively identifies the cattle mouth area during regurgitation, along with other relevant behaviors, as highlighted by the red box in the figure. UD-YOLOv5s exhibits higher detection accuracy in scenarios without interference from external factors or the presence of cattle pens, overlap, and other complexities within the monitoring area. However, there is room for improvement in accuracy when dealing with such challenging environmental conditions.

6 Conclusion

This paper presents a novel approach, UD-YOLOv5s, for cattle rumination behavior recognition by leveraging upper and lower jaw skeleton feature extraction in combination with deep learning techniques. A series of experiments were conducted to assess the effectiveness and generalization of the proposed method using a self-built cattle dataset in a consistent experimental environment. Initially, model training and ablation experiments were performed to compare and analyze the performance of the upper and lower jaw skeleton feature extraction method in recognizing rumination behavior. Subsequently, the accuracy and loss rates were evaluated under cross-subject and cross-view conditions to verify the generalization effect of the UD-YOLOv5s model. Moreover, a comprehensive comparative experiment was conducted with other rumination behavior recognition methods, namely MEAN-SHIFT, MASK-RCNN, YOLOv3, and YOLOv5s. The results demonstrate that UD-YOLOv5s outperforms these methods' recognition accuracy and performance under the same experimental conditions. In addition, it exhibits better loss rates on the cattle dataset. However, it was observed that as the size of the rumination dataset increases, the model's speed may slow down during recognition and detection due to the global detection approach implemented in the algorithm, resulting in higher computational requirements. Addressing this limitation, future research will focus on enhancing the multi-object detection

processing capabilities of UD-YOLOv5s to meet more extensive practical needs. In conclusion, this paper introduces UD-YOLOv5s, a robust cattle rumination behavior recognition method that combines skeleton extraction and deep learning techniques. The experiments and evaluations conducted on the self-built cattle dataset validate the effectiveness and generalization of the proposed method. Nonetheless, to further enhance its adaptability to complex environmental conditions, additional research efforts will be directed toward addressing challenges caused by varying lighting conditions, thus catering to a broader range of practical applications.

Informed Consent Statement

Not applicable.

Data Availability

The data used to support this work are available from the corresponding author upon request.

Acknowledgments

Conceptualization, formal analysis, methodology, resources and validation were completed by Guohong Gao, and data curation, investigation, methodology, review writing and editing was conducted by Chengchao Wang, Jianping Wang completed the data curation, resources, software, supervision, Yingying Lv conducted the Investigation and methodology, Qian Li completed the data curation and validation, Xueyan Zhang conducted the Investigation, methodology and validation, Zhiyu Li completed the data curation and supervision, Guanglan Chen completed the data curation and methodology. All authors have read and agreed to the published version of the manuscript. This work was partly supported by the Key Scientific and Technological Project of Henan Province (Grant Nos. 232102111128, 222102320181, and 212102310087), in part by the Innovation and Entrepreneurship Training Program of National College Students in China (Grant No. 202110467001), in part by the Major Special Project of Xinxiang City (Grant No. 21ZD003), and in part by the Key Scientific Research Projects of Colleges and Universities in Henan Province (Grant Nos. 23B520003 and 21A520001). The authors approved the version of the paper to be published. They agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The authors declare that there is no conflict of interest with respect to the publication of this paper. The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. A. Awad, "From classical methods to animal biometrics: a review on cattle identification and tracking," *Comput. Electron. Agric.* **123**(8), 423–435 (2016).
2. W. Andrew et al., "Automatic individual Holstein Friesian cattle identification via selective local coat pattern matching in RGB-D imagery," in *IEEE Int. Conf. Image Process.*, pp. 484–488 (2016).
3. L. E. Wallace et al., "Readability of thirteen different radio frequency identification ear tags by three different multi-panel reader systems for use in beef cattle," *Prof. Anim. Sci.* **24**(3), 384–391 (2008).
4. W. Andrew et al., "Visual localization and individual identification of Holstein Friesian cattle via deep learning," in *IEEE Int. Conf. Comput. Vis. Workshops*, pp. 2850–2859 (2017).
5. O. Guzhva et al., "Now you see me: convolutional neural network-based tracker for dairy cows," *Front. Robot. AI* **5**, 107 (2018).
6. J. Redmon et al., "You only look once: unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. and Pattern Recognit.*, pp. 779–788 (2016).
7. T. Uchino et al., "Individual identification model and method for estimating social rank among herd of dairy cows using YOLOV5," in *IEEE 20th Int. Conf. Cogn. Inf. & Cogn. Comput.*, pp. 235–241 (2021).
8. M. R. Borchers et al., "A validation of technologies monitoring dairy cow feeding, ruminating, and lying behaviors," *J. Dairy Sci.* **99**(12), 7458–7466 (2016).
9. G. Bishop-Hurley et al., "An investigation of cow feeding behavior using motion sensors," in *IEEE Int. Instrum. and Meas. Technol. Conf. (12MTC) Proc.*, pp. 1285–1290 (2014).
10. R. Arablouei et al., "*In-situ* classification of cattle behavior using accelerometry data," *Comput. Electron. Agric.* **183**(134), 106045 (2021).

11. L. J. Watt et al., "Differential rumination, intake, and enteric methane production of dairy cows in a pasture-based automatic milking system," *J. Dairy Sci.* **98**(59), 7248–7263 (2015).
12. M. Rombach et al., "Evaluation and validation of an automatic jaw movement recorder (Rumi Watch) for ingestive and rumination behaviors of dairy cows during grazing and supplementation," *J. Dairy Sci.* **101**(3), 2463–2475 (2018).
13. U. Braun et al., "Evaluation of eating and rumination behavior in cows using a noseband pressure sensor," *BMC Vet. Res.* **9**(1), 1–8 (2013).
14. P. Gregorini et al., "Rumination behavior of grazing dairy cows in response to restricted time at pasture," *Livestock Sci.* **146**(1), 95–98 (2012).
15. R. Handcock et al., "Monitoring animal behavior and environmental interactions using wireless sensor networks, GPS collars and satellite remote sensing," *Sensors* **9**(5), 3586–3603 (2009).
16. K. Schirmann et al., "Short communication: rumination and feeding behavior before and after calving in dairy cows," *J. Dairy Sci.* **96**(11), 7088–7092 (2013).
17. N. Zehner et al., "System specification and validation of a noseband pressure sensor for measurement of ruminating and eating behavior in stable-fed cows," *Comput. Electron. Agric.* **136**(29), 31–41 (2017).
18. G. M. Pereira et al., "Technical note: validation of an ear-tag accelerometer sensor to determine rumination, eating, and activity behaviors of grazing dairy cattle," *J. Dairy Sci.* **101**(3), 2492–2495 (2018).
19. S. Ruuska et al., "Validation of a pressure sensor-based system for measuring eating, rumination and drinking behaviour of dairy cattle," *Appl. Animal Behav. Sci.* **174**(3), 19–23 (2016).
20. S. M. C. Porto et al., "A computer vision-based system for the automatic detection of lying behaviour of dairy cows in free-stall barns," *Biosyst. Eng.* **115**(2), 184–194 (2013).
21. O. Yazdanbakhsh et al., "An intelligent system for livestock disease surveillance," *Inf. Sci.* **378**(15), 26–47 (2017).
22. C. Arcidiacono et al., "Moving mean-based algorithm for dairy cow's oestrus detection from uniaxial-accelerometer data acquired in a free-stall barn," *Comput. Electron. Agric.* **175**, 105498 (2020).
23. S. Viazzi et al., "Analysis of individual classification of lameness using automatic measurement of back posture in dairy cattle," *J. Dairy Sci.* **96**(1), 257–266 (2013).
24. C. Pahl et al., "Feeding characteristics and rumination time of dairy cows around estrus," *J. Dairy Sci.* **98**(1), 148–154 (2015).
25. M. Lee et al., "Wearable wireless biosensor technology for monitoring cattle: a review," *Animals* **11**(10), 2779 (2021).
26. R. Jonsson et al., "Oestrus detection in dairy cows from activity and lying data using on-line individual models," *Comput. Electron. Agric.* **76**(1), 6–15 (2011).
27. U. Braun et al., "Evaluation of eating and rumination behaviour in 300 cows of three different breeds using a noseband pressure sensor," *BMC Vet. Res.* **11**(3), 1–6 (2015).
28. B. Jiang et al., "FLYOLOv3 deep learning for key parts of dairy cow body detection," *Comput. Electron. Agric.* **166**(521), 104982 (2019).
29. P. M. Shakeel et al., "A deep learning-based cow behavior recognition scheme for improving cattle behavior modeling in smart farming," *Comput. Electron. Agric.* **19**, 10053 (2022).
30. C. Chen et al., "Behavior recognition of pigs and cattle: journey from computer vision to deep learning," *Comput. Electron. Agric.* **187**, 106255 (2021).
31. Y. Peng et al., "Dam behavior patterns in Japanese black beef cattle prior to calving: Automated detection using LSTM-RNN," *Comput. Electron. Agric.* **169**, 105178 (2020).
32. T. Tamura et al., "Dairy cattle behavior classifications based on decision tree learning using 3-axis neck-mounted accelerometers," *Animal Sci. J.* **99**(72), 589–596 (2019).
33. A. Fuentes et al., "Deep learning-based hierarchical cattle behavior recognition with spatiotemporal information," *Comput. Electron. Agric.* **177**(852), 105627 (2020).
34. J. McDonagh et al., "Detecting dairy cow behavior using vision technology," *Agriculture* **11**(7), 675 (2021).
35. R. Dutta et al., "Dynamic cattle behavioral classification using supervised ensemble classifiers," *Comput. Electron. Agric.* **111**(63), 18–28 (2015).
36. Y. Chen et al., "Intelligent monitoring method of cow ruminant behavior based on video analysis technology," *Int. J. Agric. Biol. Eng.* **10**(82), 194–202 (2017).
37. T. Li et al., "Tracking multiple target cows ruminant mouth areas using optical flow and inter-frame difference methods," *IEEE Access* **7**(198), 185520–185531 (2019).
38. T. Tamura et al., "Short communication: detection of mastication speed during rumination in cattle using 3-axis, neck-mounted accelerometers and fast Fourier transfer algorithm," *J. Dairy Sci.* **103**(8), 7180–7187 (2020).
39. D. Dutta et al., "MOOnitor: an IoT based multi-sensory intelligent device for cattle activity monitoring," *Comput. Electron. Agric.* **333**(292), 113271 (2022).
40. B. Xu et al., "Automated cattle counting using Mask R-CNN in quadcopter vision system," *Comput. Electron. Agric.* **171**(598), 100530 (2020).

41. H. Karmouni et al., "Fast computation of 3D discrete invariant moments based on 3D cuboid for 3D image classification," *Circuits Syst. Signal Process.* **40**, 3782–3812 (2021).
42. M. Yamni et al., "Accurate 2D and 3D images classification using translation and scale invariants of Meixner moments," *Multimedia Tools Appl.* **80**(17), 26683–26712 (2021).
43. M. Yamni et al., "Fast and accurate computation of 3D Charlier moment invariants for 3D image classification," *Circuits Syst. Signal Process.* **40**(12), 6193–6223 (2021).
44. O. El Ogri et al., "3D image recognition using new set of fractional-order Legendre moments and deep neural networks," *Signal Process. Image Commun.* **98**, 116410 (2021).
45. M. A. Tahiri et al., "Optimal 3D object reconstruction and classification by separable moments via the Firefly algorithm," in *Int. Conf. Intell. Syst. and Comput. Vis. (ISCV)*, pp. 1–8 (2022).
46. O. El Ogri et al., "A new fast algorithm to compute moment 3D invariants of generalized Laguerre modified by fractional-order for pattern recognition," *Multidimens. Syst. Signal Process.* **32**, 431–464 (2021).
47. H. Karmouni et al., "Fast computation of 3D Meixner's invariant moments using 3D image cuboid representation for 3D image classification," *Multimedia Tools Appl.* **79**, 29121–29144 (2020).

Biographies of the authors are not available.