

## Article

# Multi-Target Rumination Behavior Analysis Method of Cows Based on Target Detection and Optical Flow Algorithm

Ronghua Gao <sup>1,2,\*</sup>, Qihang Liu <sup>1,2,3</sup>, Qifeng Li <sup>1,2,\*</sup>, Jiangtao Ji <sup>3</sup>, Qiang Bai <sup>1,2</sup>, Kaixuan Zhao <sup>3</sup>  
and Liuyiyi Yang <sup>1,2,4</sup>

<sup>1</sup> Research Center of Information Technology, Beijing Academy of Agriculture and Forestry Sciences, Beijing 100097, China; yuqy@nercita.org.cn (Q.L.); eooz@nwafu.edu.cn (Q.B.); 202230721124@bua.edu.cn (L.Y.)

<sup>2</sup> National Engineering Research Center for Information Technology in Agriculture, Beijing 100097, China

<sup>3</sup> College of Agricultural Equipment Engineering, Henan University of Science and Technology, Luoyang 471003, China; jjtao@haust.edu.cn (J.J.)

<sup>4</sup> College of Computer and Information Engineering, Beijing University of Agriculture, Beijing 100096, China

\* Correspondence: gaorh@nercita.org.cn (R.G.); liqf@nercita.org.cn (Q.L.)

**Abstract:** Rumination behavior is closely associated with factors such as cow productivity, reproductive performance, and disease incidence. For multi-object scenarios of dairy cattle, ruminant mouth area images accounted for little characteristic information, which was first put forward using an improved Faster R-CNN target detection algorithm to improve the detection performance model for the ruminant area of dairy cattle. The primary objective is to enhance the model's performance in accurately detecting cow rumination regions. To achieve this, the dataset used in this study is annotated with both the cow head region and the mouth region. The ResNet-50-FPN network is employed to extract the cow mouth features, and the CBAM attention mechanism is incorporated to further improve the algorithm's detection accuracy. Subsequently, the object detection results are combined with optical flow information to eliminate false detections. Finally, an interpolation approach is adopted to design a frame complementary algorithm that corrects the detection frame of the cow mouth region. This interpolation algorithm is employed to rectify the detection frame of the cow's mouth region, addressing the issue of missed detections and enhancing the accuracy of ruminant mouth region detection. To overcome the challenges associated with the inaccurate extraction of small-scale optical flow information and interference between different optical flow information in multi-objective scenes, an enhanced GMFlowNet-based method for multi-objective cow ruminant optical flow analysis is proposed. To mitigate interference from other head movements, the MeanShift clustering method is utilized to compute the velocity magnitude values of each pixel in the vertical direction within the intercepted ruminant mouth region. Furthermore, the mean square difference is calculated, incorporating the concept of range interquartile, to eliminate outliers in the optical flow curve. Finally, a final filter is applied to fit the optical flow curve of the multi-object cow mouth movement, and it is able to identify rumination behavior and calculate chewing times. The efficacy, robustness, and accuracy of the proposed method are evaluated through experiments, with nine videos capturing multi-object cow chewing behavior in different settings. The experimental findings demonstrate that the enhanced Faster R-CNN algorithm achieved an 84.70% accuracy in detecting the ruminant mouth region, representing an improvement of 11.80 percentage points over the results obtained using the Faster R-CNN detection approach. Additionally, the enhanced GMFlowNet algorithm accurately identifies the ruminant behavior of all multi-objective cows, with a 97.30% accuracy in calculating the number of ruminant chewing instances, surpassing the accuracy of the FlowNet2.0 algorithm by 3.97 percentage points. This study provides technical support for intelligent monitoring and analysis of rumination behavior of dairy cows in group breeding.

**Keywords:** GMFlowNet; optical flow; regurgitation behavior; object detection; faster RCNN



**Citation:** Gao, R.; Liu, Q.; Li, Q.; Ji, J.; Bai, Q.; Zhao, K.; Yang, L. Multi-Target Rumination Behavior Analysis Method of Cows Based on Target Detection and Optical Flow Algorithm. *Sustainability* **2023**, *15*, 14015. <https://doi.org/10.3390/su151814015>

Academic Editor: George K. Symeon

Received: 25 July 2023

Revised: 11 September 2023

Accepted: 18 September 2023

Published: 21 September 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

With the advancement of modern animal husbandry, researchers are increasingly recognizing the significance of precision farming techniques, which integrate modern breeding and management practices. Precision farming entails the consideration of the individual variations among farmed animals and the development of customized farming programs. Conventional farming methods are often laborious and time-consuming, and achieving precision farming relies on a comprehensive understanding of each dairy cow [1–4].

The process of regurgitating undigested food from the esophagus to the mouth, re-chewing it, and then swallowing it is known as rumination [5–7]. Rumination mainly occurs in some herbivorous animals of the mammalian class artiodactylus, such as cattle and camels. The rumination behavior of dairy cows is closely linked to their production and reproductive performance [8–15]. Studies by Stone, Kaufman, and Antanaitis et al. have revealed a positive correlation between rumination time and milk yield (correlation coefficient = 0.30) [16–19]. Almeida et al. have demonstrated that cows with hoof disease exhibit significantly reduced rumination time; the rumination of lame cows ( $18.41 \pm 3.34$  min) was significantly reduced compared with that of healthy cows ( $28.57 \pm 3.06$  min) [20]. Hence, monitoring the individual rumination behavior of dairy cows can offer timely insights into their health status, facilitate early detection of illnesses, prevent the exacerbation of diseases impacting milk production and reproduction, and enhance dairy farming practices [21–25].

The visual characteristics of rumination behavior involve a sequence of regurgitation, chewing, and swallowing, which takes approximately 70 s to complete, with 50–60 s devoted to chewing per rumination event. Therefore, chewing frequency serves as the most perceptible indication of rumination behavior, and a decrease in chewing frequency indicates abnormal rumination. The conventional manual method of monitoring and identifying abnormal rumination behavior is inefficient and less accurate. With the expansion of farming operations and the advancement of modern precision animal husbandry, intelligent monitoring devices have emerged [26–29], enabling intelligent rumination behavior monitoring. For example, a wearable monitoring device for cow rumination behavior based on multi-source information perception and a monitoring system for cow behavior characteristics can identify cow behavior data by assembling wireless sensor nodes on the neck of cows and using algorithm analysis. Most existing intelligent rumination behavior monitoring systems employ wearable devices that track cow chewing acceleration, sound signals, and attitude information. This technology has strong accuracy and real-time performance, but the use of wearable monitoring devices will incur installation costs and equipment costs and easily causes stress reactions in cows [30].

The popularization and application of surveillance cameras in dairy farming provided the foundation for analyzing individual cow behavior using video analysis technology [31]. Researchers have utilized depth cameras, video cameras, and handheld imaging devices to capture images for evaluating cow conditions, detecting respiratory abnormalities and identifying behavior through methods such as deep learning, point cloud analysis, and video analysis. These image acquisition devices offer more comprehensive and detailed farming data compared to single information acquisition devices like wearable accelerometers, temperature sensors, and sound sensors. Moreover, they possess the advantages of being non-contact, cost-effective, and low-stress [32–35]. From the perspective of ruminant behavior monitoring, with the increasing adoption of vision technology in farming, video-image-based methods for monitoring cow rumination behaviors are emerging as a new trend for future development. These methods reduce the reliance on contact devices for monitoring cow rumination behavior and mitigate the stress caused to cows.

Chen et al. employed the Mean-Shift algorithm to accurately track the movement of a cow's jaw and extracted the center-of-mass trajectory curve of the cow's mouth movement from six videos consisting of a total of 24,000 frames for monitoring cow rumination behavior [36]. Experimental results demonstrated 92.03% accuracy for this method. Reiter et al. developed the Smartwell video analysis system using the Smartbow algorithm to monitor cow rumination time per hour [37]. They collected a sample of 100 out of

2490 single-object cow rumination videos, each lasting for 1 h, and compared the system's statistics on rumination time and the number of ruminations with manual observations. The experimental results revealed high accuracy in monitoring the time and number of cow ruminations, with an error of only 35.3 s per hour of rumination. Mao Yanru et al. utilized Kalman filtering and Hungarian algorithms to correlate and match the upper and lower jaw regions of the same cow, obtaining mouth chewing curves for achieving multi-object cow mouth tracking and rumination behavior monitoring [38]. To validate the effectiveness of the cow mouth tracking method and multi-object cow chewing behavior determination, they created a YOLOv4 model experimental dataset by collecting 66 videos of cows chewing in a real farm environment. The results indicated an average correct chewing count of 96.93%.

Video-based methods for monitoring ruminant behavior are highly sensitive to variations in lighting conditions, making them unsuitable for farming environments with significant fluctuations in light and darkness. On the other hand, due to the fact that the optical flow method is not limited by lighting conditions and has a high detection rate, it is widely used for analyzing the behavior of cows in complex breeding environments. Zhao Kaixuan et al. introduced a Horn–Schunck-based method for calculating respiration rate and detecting abnormalities in dairy cows [39]. This method employs the optical flow technique to calculate the motion velocity of each pixel in the cow breathing video image. It then applies Otsu's method to process and filter the breathing motion points in each pixel, fits the velocity direction curve, calculates the curve period, and determines abnormal cow breathing based on the duration of single breaths. Test results using a cow breathing video dataset showed an 89.06% success rate in detecting abnormal cow breathing and a 95.68% accuracy in calculating breathing frequency. Song Huaibo et al. utilized the Lucas–Kanade sparse optical flow method to extract the optical flow at the boundary of the cow's patch [40]. By computing the average optical flow direction at the patch boundary across the video sequence frames and analyzing its change pattern, they detected and analyzed cow breathing behaviors. The algorithm's frame processing time ranged from 0.10 to 0.13 s, and the average detection accuracy was 98.58%. Li Tong proposed a Horn–Schunck-based automatic detection method for the multi-objective cow mouth region [41]. They first superimposed the regions with significant optical flow changes in multiple frames to extract the cow's mouth region. Then, they combined it with the inter-frame difference method to track the mouth region and achieve intelligent monitoring of regurgitation behavior, resulting in a final detection success rate of 89.12%. Ji et al. enhanced the FlowNet2.0 optical flow method to analyze cow regurgitation behavior [42]. To eliminate interference information, they calculated the vertical component of optical flow at global pixel points and determined the chewing frequency of cows through optical flow images and curve analysis. The test results demonstrated a detection accuracy of 99.39% [42].

In conclusion, considerable research has been conducted in the field of monitoring cow regurgitation behavior, leading to significant achievements. However, the research on cow regurgitation monitoring based on machine vision technology still faces the following challenges: (1) Accurate detection of the cow's mouth area. The cow's mouth area is small and irregular, and existing object area detection methods are primarily designed for objects with more distinct characteristics, making it difficult to effectively detect the cow's mouth area. (2) Although the rumination behavior monitoring method based on video analysis has good effects, it has strong light sensitivity and is not suitable for the breeding environment with large changes in light and shade. (3) Monitoring cow rumination behavior in multi-object scenarios. The farm environment is complex, and cows typically live in herds, requiring the simultaneous monitoring of multiple cows. However, most existing rumination monitoring methods focus on single cows, resulting in a lower efficiency.

To address these challenges, this study focuses on the precise breeding requirements for the intelligent monitoring of cow regurgitation behavior in barn groups. It investigates precise detection methods for multi-object cow regurgitation areas and analyzes regurgitation behavior. In order to tackle the issues related to inaccurate extraction of small-scale

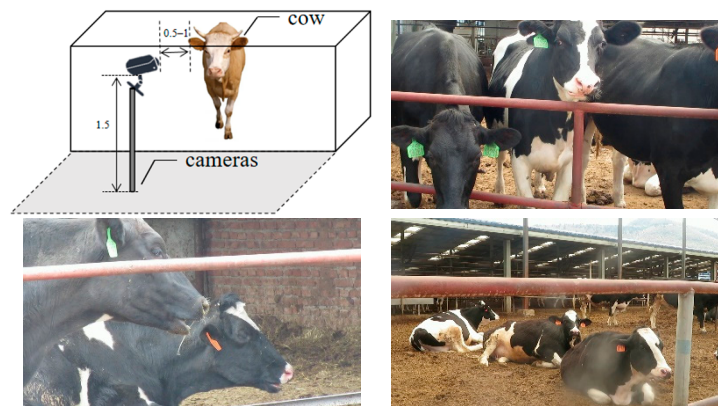
optical flow information and interference between different optical flow information in multi-object scenarios, the study employs an optical flow algorithm to obtain optical flow information of the cow's mouth region in consecutive frames based on object detection results. Subsequently, the lower half of the optical flow in the cow's mouth region is extracted to calculate the perpendicularity of each pixel. By extracting the velocity magnitude values in the vertical direction from the lower half of the cow's mouth region, the interference caused by the cow's head movement, apart from the mouth movement, is mitigated. This research is crucial for realizing large-scale breeding and developing a non-contact intelligent image detection system. Furthermore, it provides technical support for intelligent monitoring and an analysis of cow regurgitation behavior in group breeding settings.

## 2. Materials and Methods

### 2.1. Data Sources

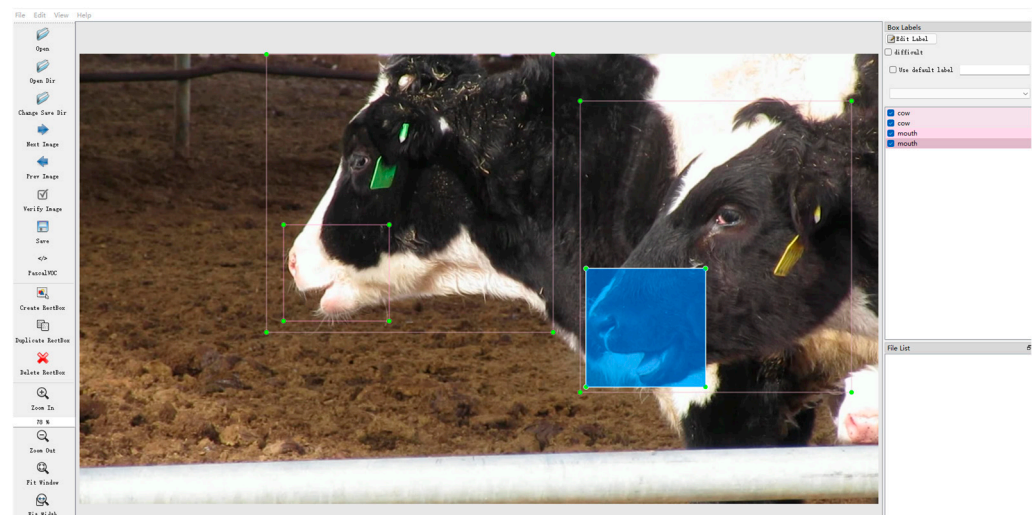
The Yanqing Dadi Qunsheng Dairy Farming Base currently has 1100 dairy cows with an annual output of about 5500. Between September 2021 and February 2023, video data capturing cow ruminant behavior were gathered from two locations: the Yanqing Dadi Qunsheng Dairy Farming Base and the Beijing Furnong Xingmu Dairy Farming Centre in Beijing. The cows collected in the dataset in this paper were Holstein cows that were about 19 months old. In order to meet the high-definition needs of shooting, HD1080 resolution industrial cameras (camera size  $105.0 \times 60.9 \times 41.4$  mm) were positioned at an angle of  $45^\circ$  to  $90^\circ$  relative to the side of the ruminating cow's head and placed at a distance of approximately 0.5–1 m from the targeted cow. The camera was placed at a height of 1.5 m from the ground because this height corresponds to the height of the mouth of the standing cow when it ruminates, which can better capture the rumination movement of the cow. The reason why the distance between the camera and the target cow is 0.5–1 m is that the recognition accuracy of the mouth area of the cow will be low if the distance is too far from the cow, and the rumination movement information cannot be accurately calculated. If the camera is too close to the cow, the cow will be affected by the camera, and the target of the cow will be captured less with the camera, then the rumination behavior of the multi-target cow cannot be analyzed. Considering the continuity of the video, we set the duration of each video to 1–2 min and the frame range from 1800 to 3600. When the location of the camera was fixed at the height mentioned in this paper and the distance from the target of the cow, the recognition accuracy of the ruminant beak region of the cow is high, and the angle influence is small. Both the side view and the front view of the lens can accurately capture the mouth movement information of the target of the cow. If it is put into actual production, the target cow can be photographed with a higher definition camera, and the camera erection angle does not affect the algorithm accuracy, so the cow target can be photographed from a distance in the cow shed.

At first, the video data pertaining to ruminant behavior were pre-processed. This involved segmenting approximately 10 s of clear video footage, comprising 300 frames per segment (Figure 1). In these segments, the cow's head dominated the camera's field of view, and the mouth exhibited prominent chewing movements during rumination. However, the ruminant process was occasionally accompanied by slight random head oscillations, which impeded the extraction of ruminant optical flow features due to their interference with the movements of the neighboring cows.



**Figure 1.** Acquisition diagram and video example of cow rumination.

Randomly selected key frames from the cow regurgitation videos were extracted, and the cow object and mouth region in these frames were annotated using the Labelling software (App version: 1.8.6), as depicted in Figure 2. A total of 5939 cow regurgitation images were labeled to construct a training dataset for multi-objective cow regurgitation analysis.

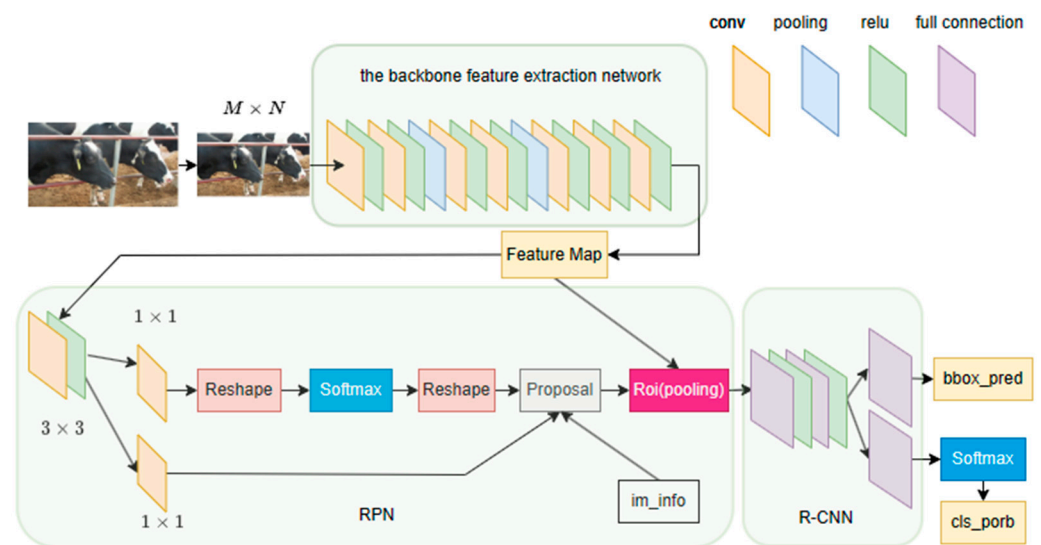


**Figure 2.** Multi-objective cow ruminant dataset labeling.

## 2.2. Theoretical Approach

The Faster R-CNN [43] two-stage algorithm offers advantages over single-stage object detection algorithms due to its ability to extract object features more accurately, resulting in higher accuracy. Therefore, this study enhances the Faster R-CNN as the foundational object detection network.

The architecture of the Faster R-CNN network comprises three components: the backbone feature extraction network, the candidate region extraction network (RPN), and the object detection network (regions with CNN Features (R-CNN)). The algorithm's structure is illustrated in Figure 3.



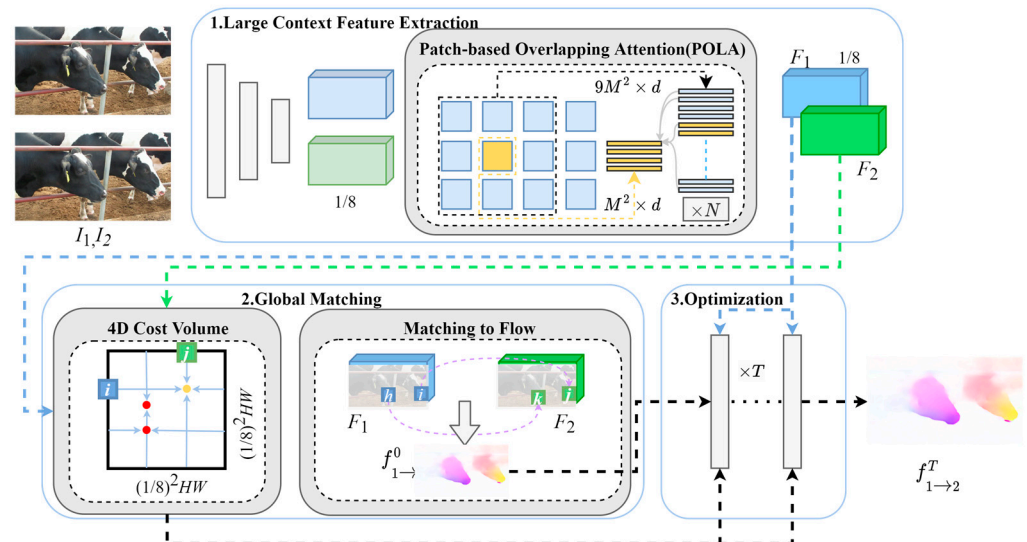
**Figure 3.** Structure diagram of Faster R-CNN.

Initially, the algorithm rescales the input image to a specific proportion. The VGG16 backbone feature extraction network is then utilized to extract features from the head region and mouth region of the rescaled image. Through successive convolution, pooling, and full concatenation operations, a feature map corresponding to the original image is generated. This output result is shared with the RPN network. The RPN network slides windows across the feature map, generating a specific set of anchor points for each window. By convolving these anchor points with the feature map, the network splits the feature map with a convolution of  $3 \times 3$ . The resultant initial prediction frames are aligned with the feature map obtained from the VGG16 backbone feature network in the previous step. Traversing the image leads to the generation of numerous rectangles of various sizes around the original image object. Next, the algorithm applies further filtering to the initial prediction frames by employing a fixed threshold, yielding candidate frames. A fixed threshold convolution of  $1 \times 1$  is set to adjust the vertex coordinates of the candidate frame in the original image. A sequence of correction operations is performed on the adjusted frame to obtain the prediction frame for the object to be detected. Subsequently, the object detection network R-CNN aligns the prediction frame generated using the candidate region extraction network with the result obtained from the backbone feature extraction network. This alignment process enables the determination of the region of interest (ROI). The ROI undergoes full concatenation and SoftMax classification operations, producing classification confidence data and object prediction frame coordinates.

Detecting the mouth regions of multi-object cows poses a challenge due to the small size of the cow's mouth region, leading to low accuracy with the Faster R-CNN algorithm. To address these issues, this study proposes improvements to the Faster R-CNN algorithm to enhance its accuracy in detecting the mouth region of multi-object cows.

Given the slight displacement associated with cow regurgitation behavior, it is necessary to employ a highly accurate global optical flow analysis algorithm to capture the characteristic patterns of regurgitation behavior. The complexity and variability of cow regurgitation optical flow in multi-objective scenarios, along with the need for more precise optical flow estimation for smaller-scale objects, call for an improved optical flow algorithm. Existing optical flow estimation methods based on direct regression using neural networks struggle to capture long-term movement correlations, resulting in an inaccurate acquisition of multi-object regurgitation optical flow information. To enhance the accuracy of optical flow estimation, this study introduces the GMFlowNet [44] optical flow algorithm, which incorporates global matching before direct regression and utilizes the POLA attention mechanism for feature extraction. These enhancements make GMFlowNet well-suited for optical flow estimation in regions with large motion and textures. Consequently, in this

study, GMFlowNet is adopted as the underlying optical flow estimation network to acquire more accurate, smaller-scale multi-object optical flow information. The network structure of GMFlowNet is illustrated in Figure 4.



**Figure 4.** GMFlowNet network structure.

GMFlowNet comprises three components: a feature extraction network, global matching, and learning-based optimization. The feature extraction network is responsible for extracting relevant feature information from the input image sequence. The GMFlowNet algorithm incorporates Large Context Feature Extraction to address the challenge of local fuzzy matching at key locations. Initially, features are extracted using three convolutional layers. Transformer blocks are then employed to handle long-term dependency information. Patch-based Overlapping Attention (POLA) is applied to extract large context features, reducing regional ambiguity in matching. Unlike Swin Transformer's sliding window approach that necessitates two separate attention blocks for exchanging information between patches, which leads to information loss and is not conducive to matching, POLA incorporates features between patches within a single block, facilitating direct information exchange. This results in improved performance with reduced information loss and lower memory consumption. The global matching process consists of a 4D cost calculation and matching confidence computation. A 4D cost volume is constructed based on the input image pair at a resolution of 1/8. The cost volume is then transformed into matching confidence scores using the double SoftMax operator. The optical flow is optimized with the learning-based optimization reference RAFT, changing the initial RAFT optimization parameter from 0 to  $f^0_{1 \rightarrow 2}$ . GMFlowNet introduces a more lightweight network structure that reduces computational effort and model size while maintaining higher accuracy. Compared to other algorithms, GMFlowNet can process larger image sizes, resulting in more precise optical flow information, thanks to its lower video memory requirements.

However, the high accuracy of GMFlowNet makes it vulnerable to motion interference from sources other than cow rumination when analyzing the optical flow of rumination motion in complex scenes. To address this challenge, this study enhances the Faster R-CNN algorithm to extract the mouth regions of multiple objects, thereby rejecting background optical flow interference. Additionally, improvements are made to the GMFlowNet algorithm to mitigate interference from other head movements, further enhancing its accuracy in the task of analyzing multi-object cow regurgitation optical flow.

### 3. Improved Faster R-CNN Object Detection Algorithms

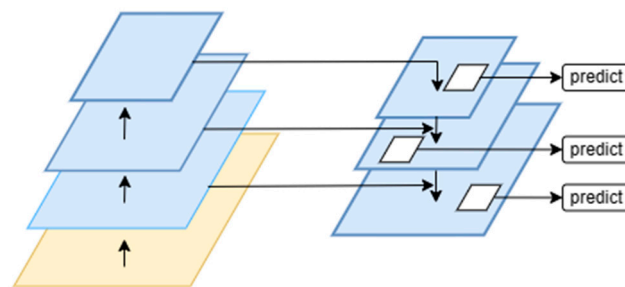
The dataset in this paper contains cow targets in different scenes (night, day, rest area, active area, and feeding area) and from different angles (side view, front view, and down

view), so the algorithm has a high recognition accuracy in different angles and positions. To enhance the detection performance of Faster R-CNN in identifying the cow's mouth region, this research incorporates the Convolutional Block Attention Module (CBAM [45]) attention mechanism into the Faster R-CNN backbone network. Additionally, it integrates the feature pyramid network (FPN [45]) and utilizes the ResNet-50-FPN network to extract more effective features from the cow's mouth region, thereby improving the model's ability to detect the cow's mouth region accurately. Moreover, the integration of optical flow information helps eliminate false detection objects. Furthermore, an interpolation approach is employed to supplement and refine the detection frame of the cow's mouth region, ensuring that no detections are missed and enhancing the overall accuracy of cow mouth ruminant region detection.

### 3.1. Fusion Feature Pyramid Network

Considering that cows are herd animals and are typically bred in large and complex environments, industrial cameras capture multi-object video images of cow regurgitation behavior from various camera positions, resulting in significant variations in the proximity and distance scales between individual cows. Moreover, the image features in the mouth region of cows differ depending on their scales. Therefore, it is crucial to employ a method capable of extracting features from the mouth region of cows at different scales to ensure accurate detection.

The original feature map possesses a high resolution, enabling more precise and detailed location information for each region of the object cow. Consequently, it is more suitable for detecting small-scale objects. Conversely, the scaled feature map accentuates the differences between objects in the image, making it more appropriate for detecting large-scale objects. However, the original Faster R-CNN model exclusively employs the scaled feature map for detecting the cow's mouth region, resulting in a lack of detailed information and significant issues related to missed detection of smaller-scale cow mouth regions. The overall structure of the FPN network is illustrated in Figure 5.



**Figure 5.** Structure diagram of FPN.

The network architecture consists of three key components: the bottom-up path, the top-down path, and the lateral connections connecting them. The bottom-up path processes the pre-processed image through a feature extraction network, similar to a CNN convolution operation, using a forward propagation approach. The Faster R-CNN network classifies features based on different feature map sizes, ensuring that the scale of feature maps for adjacent categories remains relatively consistent. Each category corresponds to a level within the feature pyramid, serving as a basis for classifying features at different levels in the FPN. The top-down path involves scaling down the smallest-scale feature map to match the size of the feature map at the previous level through upsampling. This strategy enables the localization of feature information for classification and facilitates subsequent feature extraction. The method leverages the smallest-scale feature map, which contains more pronounced semantic information, for feature localization of large-scale objects. Simultaneously, it utilizes the detailed information within the largest-scale feature map to capture feature information of smaller-scale objects.



FPN employs a lateral fusion structure to integrate semantic information with image detail information. The lateral fusion process ensures dimensional consistency across different classes of feature maps and enables the comprehensive fusion of feature information across various scales.

### 3.2. CBAM Attention Mechanism

In order to enhance the accuracy of feature extraction, the fused feature maps undergo additional processing using the Convolutional Block Attention Module (CBAM) attention mechanism. In this study, the Convolutional Block Attention Module (CBAM) attention mechanism is added to the backbone feature extraction network. This mechanism assigns weights to feature information based on its contribution to the detection results. Unlike a single max-pooling or averaging operation, the attention module employs a combination of max-pooling and averaging in an additive or stacked manner. The Spatial Attention Module (SAM) is no longer limited to a single pooling or averaging operation but rather a summation or stacking of pooling and averaging operations. The structure of this module is depicted in Figure 6.

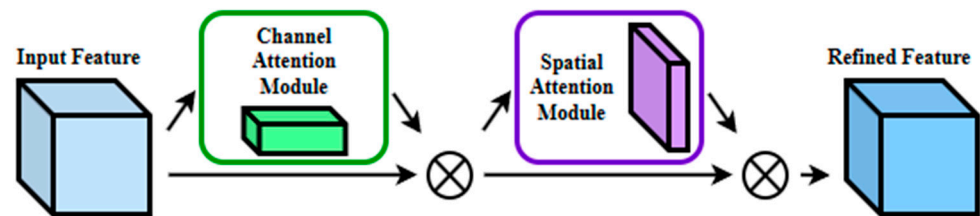


Figure 6. Structure diagram of CBAM.

The channel attention module directs attention to the essential content of the target object. It achieves this by applying global average pooling and global maximum pooling operations to the incoming feature maps. The results of these pooling operations are then processed using a shared fully connected layer. Finally, the weights obtained from this processing are combined with the input feature layer through a weighting calculation. This is calculated using

$$M_c(F') = \sigma(\text{MLP}(\text{AvgPool}(F) + \text{MaxPool}(F))) \quad (1)$$

where  $M_c(F')$ —channel attention weights and  $\sigma$ —Sigmoid function. MLP—Multi-Layer Perceptron and AvgPool—global average pooling. MaxPool—maximum pooling and  $F$ —the input feature map.

To compute channel attention efficiently, the CAM employs a dual pooling approach that compresses the input feature map in the spatial dimension. Unlike a single pooling method, this approach utilizes both average pooling and maximum pooling, resulting in stronger representational power.

On the other hand, the spatial attention module (SAM) is particularly sensitive to the positional information of the object within the cow's mouth region that needs to be detected. It effectively reduces the interference from background information. To calculate spatial attention, average pooling and maximum pooling are first applied along the channel direction of different feature points and stacked together to generate a comprehensive feature descriptor. This descriptor consists of two distinct 2D maps: the average pooling feature and the maximum pooling feature for each channel. Subsequently, a standard convolution layer is utilized to concatenate and convolve these features. The Sigmoid activation function is applied to obtain a 2D spatial attention map, representing the weight assigned to each feature point in the input feature map. Finally, the channel weights are multiplied with the input feature layer. This is calculated using

$$M_s(F'') = \sigma\left(f^{7 \times 7}([\text{AvgPool}(F'); \text{MaxPool}(F')])\right) \quad (2)$$

where  $M_s(F'')$ —spatial attention weights and  $f^{7 \times 7}$ —convolution kernel size is  $7 \times 7$ .  $F'$ —feature map after channel attention weighting.

### 3.3. Fusing Optical Flow Information

The cow's mouth region occupies a relatively small portion, approximately 1%, of the entire image, making it challenging to extract sufficient information and accurately identify this category. Despite the improved accuracy of the Faster R-CNN object detection algorithm, it still encounters difficulties in continuous video frame detection, resulting in missed and erroneous detections. To ensure data accuracy, the analysis of mouth movement information for multi-object cows using the optical flow method requires a continuous input image sequence. Hence, this study combines the FlowNet 2.0 optical flow algorithm with the improved Faster R-CNN object detection results to enhance filtering and supplementation.

One of the challenges with the Faster R-CNN object detection algorithm is its sensitivity to smaller regions, leading to the misidentification of background patches as the cow's mouth in the image. Consequently, the problem of misidentification in background areas for object detection results arises. To address this issue, this research incorporates the optical flow information of the detection area to filter and eliminate incorrectly detected object frames. The specific filtering method is

$$box = \begin{cases} 0, & \mu_1 < \mu' \\ 1, & \mu_1 \geq \mu' \end{cases} \quad (3)$$

where  $box$ —cow object detection frame.  $\mu_1$ —mean value of optical flow in the detection frame area.  $\mu'$ —optical flow threshold.

At first, the multi-object detection frame in each frame is traversed as per the video frame sequence  $box$ , and the mean value of the optical flow in each detection frame area of the frame is calculated,  $\mu_1$ . This value is compared with the set threshold value  $\mu'$  to determine whether there is a lot of motion information in the area. It filters out the wrong detection object frame by rejecting it. There are two purposes of the multi-object cow object detection frame: firstly, to ensure the accuracy of the mouth region detection. If the cow object detection frame corresponds to the cow's mouth detection frame, the mouth region can be successfully identified. Secondly, it lays the foundation for the subsequent identification of individual cow information.

Rejecting the wrong frame ensures the accuracy of the detection of the cow's object and mouth region. However, the model faces the challenge of missed detection in the continuous frame detection task, primarily due to the small size of the cow's mouth region and its susceptibility to factors such as the shooting angle. To address the issue of missed detection and maintain the continuity of the detection frame in the video frame sequence, this paper implements the frame complementary algorithm using the interpolation idea.

The object detection video frame  $f$  is first traversed, the key frame in the video is selected as the detection reference frame  $f_k$ , and the number of cow objects in the video segment  $x$  is counted. The parameter  $mouth\_box$  ensures no loss of mouth optical flow information, and the  $x$  cow mouth detection frame in the  $f$  frame is expanded by an equal number to ensure that the detection frame covers the cow's jaw movement area. All mouth frames corresponding to the cow's object frame are then placed in the array  $mouth\_list$  and grouped by the number of cow objects. The first traversal process is depicted in Figure 7.

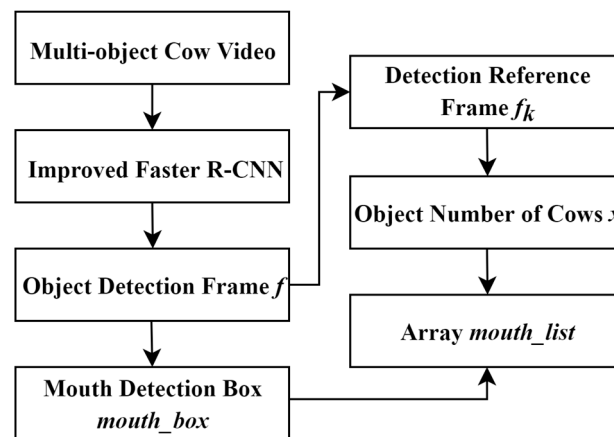


Figure 7. Flow chart of the first traversal.

After obtaining *mouth\_list*, the video frame is traversed again, and the IoU threshold of the detection frame  $\mu_{IoU}$  is set for the current detection frame  $f_x$  and the reference frame  $f_k$ . Following the order of the video frames, the frames with more than the threshold, i.e., a high degree of overlap, are filled into the corresponding cow group in *mouth\_list*. If the current frame *mouth\_box* is successfully filled in during the traversal process, the frame is replaced with a new reference frame to track the mouth region. If the detection frame of a cow object in the current frame is missing, or the IoU is less than the set threshold, the reference frame of the cow  $f_k$  is filled in the construct the frame for the missed object. The second traversal process is shown in Figure 8.

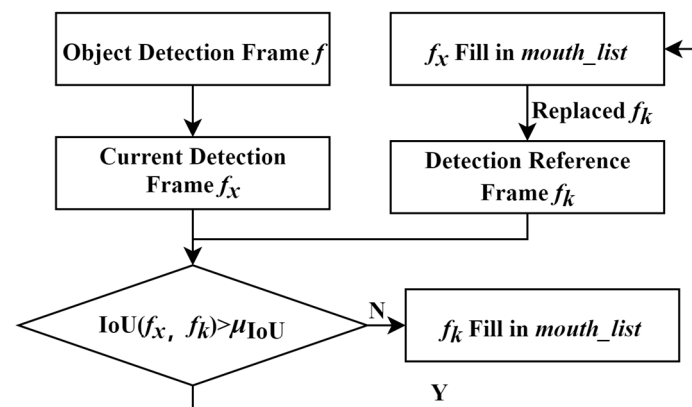


Figure 8. Flow chart of the second traversal.

#### 4. GMFlowNet-Based Multi-Objective Analysis of Cow Ruminant Behavior

The improved and optimized Faster R-CNN detection algorithm provides the basis for analyzing cow regurgitation behavior in multi-object scenarios by detecting the multi-object cow mouth regions. To address the issue of inaccurate extraction of multi-object cow regurgitation optical flow in continuous frames, this study enhances the GMFlowNet optical flow algorithm. It focuses on extracting the vertical component of optical flow velocity in the ROI region of the cow's mouth and incorporates the Mean-Shift clustering method for improved accuracy and efficiency in estimating optical flow information at mid- and far ranges. The resulting optical flow change curve reflects the movement of the multi-object cow's mouth. The Inter-Quartile Range (IQR) is used to calculate the mean square deviation and remove outliers. Subsequently, the ruminant behavior is determined, and the number of cows engaged in regurgitation chewing is calculated.

#### 4.1. Multi-Object Cow Ruminant Optical Flow Extraction

The information on cow regurgitation optical flow in multi-object scenarios is characterized by its complexity and variability. Accurately acquiring optical flow information at smaller scales requires a more precise optical flow algorithm. Neural-network-based optical flow estimation methods that rely on direct regression struggle to capture long-term movement correlations, leading to an inaccurate acquisition of multi-object regurgitation optical flow information. To overcome this, our proposed approach utilizes the improved Faster R-CNN algorithm to obtain the coordinates of the multi-object ruminant regions. GMFlowNet is then employed to extract optical flow within multiple cow mouth ROI regions, eliminating the horizontal velocity component and mitigating interference caused by horizontal motion.

The ruminant area box is labeled based on the connected area of the cow's nose and mouth to enhance object feature extraction during detection. However, a direct optical flow analysis of the object detection results would be affected by the movement of the nasal region, i.e., the cow's head. Therefore, in this section, we reduce the detection frame of the regurgitated region obtained from the improved Faster R-CNN, focusing solely on the lower half of the mouth region. The effectiveness of this interception is demonstrated in Figure 9.



**Figure 9.** Example of mouth area interception. (a) Original object detection results. (b) Mouth area interception results.

As can be seen from the figure above, the mouth detection frames of the two cows in Figure 9a are too large, with a significant amount of background area and including the cow's nose area. On the other hand, the intercepted mouth detection frame in Figure 9b contains only the mouth object area with a minimal amount of background. This reduction helps eliminate some interference factors and ensures the accuracy of the subsequent optical flow analysis.

The GMFlowNet algorithm was employed to extract optical flow data from the mouth regions of multiple objects within a single frame. To mitigate the impact of disruptive head movements of cows, the extraction of optical flow from multi-object mouth regions is converted into individual extractions for each object's mouth region. Additionally, instead of utilizing the optical flow velocity value as employed in the original algorithm, the vertical velocity component of the optical flow is calculated at each pixel point. This adjustment helps minimize the influence of disturbing head movements. The horizontal optical flow component  $u$  and the vertical optical flow component  $v$  are first calculated for all pixel points between the two video frames, and the optical flow velocity magnitude  $s$  for each pixel point between the two frames is

$$s = \sqrt{u^2 + v^2} \quad (4)$$

The angle of motion of the optical flow at each pixel  $a$  is

$$a = \frac{180 \cdot \arctan2(v, u)}{\pi} \quad (5)$$

For  $a$ , different colors were determined to represent different angles of movement, and  $s$  determined the intensity of the colors to generate an optical flow diagram of the cow's ruminant behavior.

This paper defines a complete ruminant chewing movement as a vertical motion moving up and down, starting from the cow's mouth being closed, then opening to the maximum, and finally closing again.

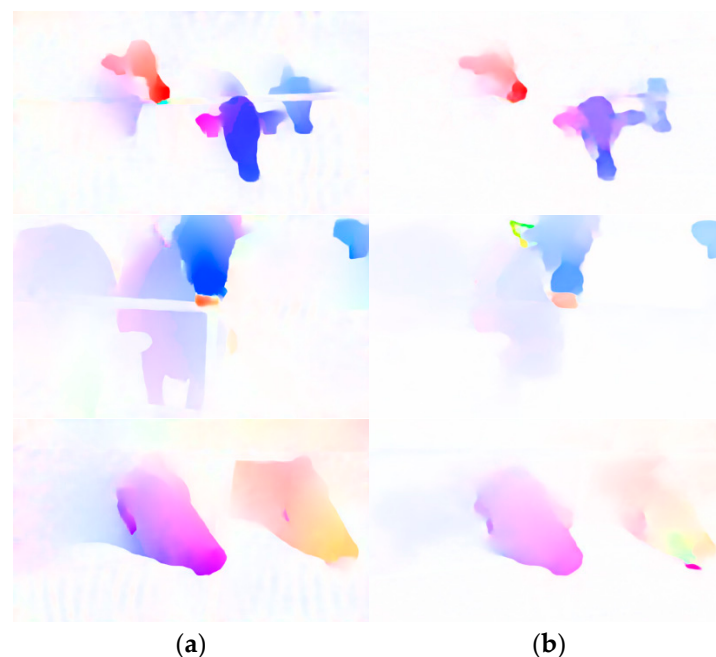
In order to avoid the influence of the background and its own motion on the optical flow analysis, the optical flow of ruminant behavior is determined from the motion components in both directions. We define  $T$  as the optical flow image and use the GMFlowNet algorithm to calculate the optical flow image  $T$ , which is given using

$$T = a + s \quad (6)$$

To reduce the influence of the horizontal motion component  $u$  in  $s$  on the extraction of the ruminant behavior features, the optical flow velocity  $s$  is replaced by the absolute value of the vertical optical flow component  $v$  for each pixel point  $|v|$  and calculated as

$$T = a + |v| \quad (7)$$

The optical flow diagrams generated before and after the improvement of the GMFlowNet algorithm are compared, as shown in Figure 10.



**Figure 10.** Comparison of optical flow diagrams with multi-objective cows. (a) GMFlowNet optical flow diagram. (b) Improved optical flow diagram.

Different colors in the figure represent different optical flow motion directions, that is, the motion directions of pixels in the video, and the color intensity of pixels is used to represent the motion speed of the point. The observation from Figure 10 reveals that within various multi-object scenes captured in the same video frames, Figure 10a predominantly shows optical flow regions attributed to cow head movements or limb movements. However, the enhanced algorithm successfully suppresses optical flow generated by interfering

movements, thereby rendering the regurgitated optical flow in the mouth region more prominent. This enhancement greatly facilitates optical flow analysis.

#### 4.2. Improved GMFlowNet Algorithm for Computing Multi-Object Optical Flow

Upon precise extraction of optical flow data from the multi-object ruminant region, the Mean-Shift clustering algorithm is integrated to cluster the vertical velocity components of each pixel within the ROI region. The algorithm utilizes the more abundant classes to compute the mean value, serving as the ruminant optical flow value for the corresponding cow object in the current frame. This approach mitigates the impact of the cow's own head movements and background interference on the extraction of ruminant optical flow information. A traversal method is employed to capture the ruminant movement patterns of the multi-object cows throughout the video frames. The traversal process involves calculating the mean square difference, incorporating the concept of quartile spacing to eliminate curve outliers, and subsequently filtering and fitting the data to determine the peaks and valleys of the curves. Through this process, it becomes possible to ascertain whether the cows are ruminating, ultimately providing accurate ruminant chewing times for the multi-object ruminant cows.

When multiple cows are regurgitating simultaneously, the optical flow in each regurgitated area becomes exceptionally complex and susceptible to interference from other movements. Even after removing the horizontal component of motion at each pixel, the optical flow in each frame remains intricate and exhibits rapid fluctuations over the time axis. Consequently, directly using the average value of optical flow at each pixel for analysis proves challenging in obtaining the optical flow pattern across the frame sequence. In order to achieve a more precise calculation of optical flow information and analyze the variation patterns, this study incorporates the Mean-Shift clustering algorithm to cluster the optical flow data at each pixel point.

The Mean-Shift algorithm identifies clusters of points within the feature space of sample points by moving toward regions of higher density. For example, in a feature space with  $N$  sample points, at first, a random centroid *center* is determined, and the set of all points in a circular space of radius  $D$  named  $x_i$ , the vector of  $x_i$ , and the centroid *center* are calculated. Then, the mean value of all vectors in the entire circular space, i.e., the offset mean of the circular space  $M$  is calculated. Finally, the centroid *center* is moved to the offset mean  $M$ , and the above process is repeated until a certain condition is met. The basic formula for Mean-Shift is

$$M(x) = \frac{1}{k} \sum_{x_i \in S_h} (x_i - x) \quad (8)$$

where  $S_h$ —a high-dimensional spherical region with  $x$  as the center point and a radius of  $h$ .  $k$ —number of points included in the range  $S_h$ .  $x_i$ —points included in the range  $S_h$ .

The formula for moving the center point to the offset mean position during the cycle is

$$x^{t+1} = M^t + x^t \quad (9)$$

where  $M^t$ —the mean value of the offset obtained in the  $t$  state.  $x^t$ —the center in the  $t$  state.

In this study, the Mean-Shift algorithm is enhanced by incorporating a kernel function that bypasses the mapping relationship and performs direct calculations within the low-dimensional space. This approach aims to transform the low-dimensional indistinguishable data into a divisible high-dimensional representation. By introducing the kernel function, the Mean-Shift algorithm assigns a higher weight to points near the center of the calculation, effectively incorporating a Gaussian weight. This weighting scheme reflects that shorter distances between points correspond to higher weights. The improved offset mean value  $M_h$  is calculated as

$$M_h(x) = \frac{\sum_{i=1}^n x_i g\left(\left|\frac{x-x_i}{h}\right|^2\right)}{\sum_{i=1}^n g\left(\left|\frac{x-x_i}{h}\right|^2\right)} - x \quad (10)$$

where  $n$ —the number of points in the bandwidth range.  $g(x)$ —the derivative of the kernel function.

The core process of the Mean-Shift algorithm involves calculating the centroid's drift vector based on the variations in data density within the region of interest (ROI) for the cow's mouth. This centroid is then moved to the next iteration until the maximum density is attained. This operation is applied to each data point while also keeping track of the number of occurrences within the ROI. The count of occurrences serves as the basis for classification. Utilizing mean values facilitates automatic determination of the number of categories without the need for manual input, thus offering an advantageous feature of the Mean-Shift algorithm.

The Mean-Shift algorithm clusters the vertical components of the optical flow at each pixel  $v$  in the optical flow data matrix of the cow's ruminant region  $mouth\_box$  and arranges the clusters in descending order of the number of elements in the set  $Q$ . The calculation process is as follows:

$$mouth\_flow = \begin{cases} C_{Q_{max}} & 0 < Q_n \leq \mu_q \\ \frac{2\sum_{i=1}^{Q_n} C_{Q_i}}{Q_n}, & Q_n > \mu_q \end{cases} \quad (11)$$

where  $mouth\_flow$ —regurgitated regional optical flow values.  $Q_n$ —number of clustered sets.  $\mu_q$ —collection number threshold.  $C$ —collection centroid value.  $Q_{max}$ —collection with the highest number of elements.  $Q_n Q_{max} C_{Q_{max}}$ . If the number of clustered sets  $Q_n$  is high, it means that the optical flow in the regurgitated region is complex, and there is more interference information—the optical flow value.

The aforementioned process is iterated across a sequence of video frames to derive the mouth movement curve for each cow object. To maintain curve continuity, the mean squared deviation is computed, and anomalous optical flow values are substituted using the concept of quartile spacing. This ensures the smooth and uninterrupted flow of the curve. The principle of  $3\sigma$  is used as a criterion for determining outliers: a set of measured values that deviate from the mean by more than three times the standard deviation is defined as an outlier. The calculation process for outlier replacement is performed as follows:

$$F_c(x) = \begin{cases} mouth\_flow, & |x - \bar{x}| \leq 3\sigma \\ Q^U, & x > 3\sigma + \bar{x} \\ Q^L, & x < \bar{x} - 3\sigma \end{cases} \quad (12)$$

where  $F_c(x)$ —mouth optical flow curve.  $\sigma$ —standard deviation.  $Q^U$ —upper quartile.  $Q^L$ —lower quartile.  $\bar{x}$ —mean value of optical flow data.

The mean value of the optical flow curve  $\bar{x}$  and the standard deviation  $\sigma$  are first calculated, then the curve is traversed, and if abnormal data are found, the abnormal data are replaced using the IQR idea. The lower quartile of  $Q^L$  indicates that one quarter of the curve is smaller than  $Q^L$ , and  $Q^L$  is used to replace the very large outliers. The upper quartile of  $Q^U$  indicates that one quarter of the curve is larger than  $Q^U$ , and  $Q^U$  is used to replace the very small outliers; the outlier replacement is achieved.

The video frame sequence serves as the horizontal coordinate, while the regurgitated motion curve of the optical flow data, which is normalized to mitigate noise from other movements of the cow's head, is represented as the vertical coordinate. To achieve this noise reduction, a Savitzky–Golay filter is employed. As the filtered ruminant movement curve shows sinusoidally distributed peaks and troughs, the fitting function  $y$  is defined and the equation is

$$y = a_0 \sin(a_1 x + a_2) + a_3 \quad (13)$$

where  $a_0$ ,  $a_1$ ,  $a_2$ , and  $a_3$ —the fitted variables.

The maximum value point of the fit function  $h$  can be calculated using the second order derivative, traversing the curve to count the number of maximum value points  $h$ ,

which is the number of times the cow chews the input video  $j$ , so that  $t$  is the time of the input video, and the number of times the cow chews  $p$  is calculated as

$$p = \frac{j}{t} \quad (14)$$

where  $j$ —number of times cows ruminate.  $t$ —input video time.

## 5. Results and Analysis

### 5.1. Experimental Dataset and Parameter Settings

The experiments were conducted on an Ubuntu 16.04 operating system, utilizing the PyTorch framework with Python version 3.8, PyTorch version 1.7.0, CUDA API version 10.0.130, and cuDNN version 7.4.1. For the experimental dataset, nine multi-object cow regurgitation videos with diverse scenes were randomly selected.

### 5.2. Experimental Analysis

In order to validate the accuracy of the enhanced Faster R-CNN algorithm in detecting the mouth region of cow regurgitation, a comparison was conducted between the original and improved models using nine multi-object cow regurgitation video frame datasets captured under consistent equipment and algorithmic conditions. The evaluation of the recognition model's performance before and after the improvement included considerations of model weight size, accuracy, completion rate, and mAP. The experimental findings are presented in Table 1.

**Table 1.** Model experiment results.

Model	Check Accuracy	Search Completeness Rate	mAP@0.5:0.95	Model Size/M
Faster R-CNN	0.8381	0.7974	0.7090	467.99
Yolov3-tiny	0.9550	0.9270	0.6020	24.30
Ours	0.9362	0.8601	0.7833	468.03

The improved Faster R-CNN model exhibits a smaller difference in model weight size compared to the original algorithm, along with a verification accuracy rate of 0.9362 and a complete verification rate of 0.8601. Furthermore, the mAP@0.5:0.95 score of 0.7833 surpasses that of the original network model, indicating a 7.43 percentage point improvement over the original algorithm. The model recall rate and precision of Faster R-CNN are slightly lower than that of Yolov3-tiny, but mAP@0.5:0.95 is 18.13 percentage points higher than that of Yolov3-tiny.

To validate the accuracy of the enhanced Faster R-CNN algorithm in detecting the regurgitated mouth region of cows, nine datasets consisting of video frames from multi-objective cow regurgitation scenarios were selected under identical equipment and algorithm environments. The experimental comparison of regurgitated region accuracy between the Faster R-CNN algorithm and the improved algorithm is presented in Table 2.

As indicated in Table 2, the Faster R-CNN algorithm exhibited an average ruminant mouth region detection accuracy of 72.90%. However, the improved Faster R-CNN algorithm achieved an average ruminant mouth region detection accuracy of 84.70%, surpassing the previous accuracy by 11.80 percentage points.

Videos 1–3 encompassed long-range multi-object regurgitation data, where the cow's mouth area was relatively small, resulting in reduced feature information and consequently lower detection accuracy. In contrast, video 11 consisted of medium- to close-range multi-object regurgitation data, where the detection success rate for the mouth area was higher. A selection of randomly chosen frames from the detection images was compared between



the original Faster R-CNN algorithm and the enhanced algorithm, with the experimental results presented in Figure 11.

**Table 2.** Comparative analysis of different methods for ruminant area accuracy in dairy cows.

Video Serial Number	Number of Actual Regurgitated Areas	Value Calculated with the Algorithm before Improvement	Regurgitation Area Accuracy/%	Improved Algorithm Calculated Values	Regurgitation Area Accuracy/%
1	1500	639	42.60	1017	67.80
2	900	306	34.00	561	62.33
3	900	540	60.00	628	69.78
4	600	596	99.33	598	99.67
5	330	316	95.76	320	96.97
6	600	494	82.33	567	94.50
7	600	510	85.00	589	98.17
8	900	691	76.78	727	80.78
9	600	482	80.33	554	92.33
Average	-	-	72.90	-	84.70

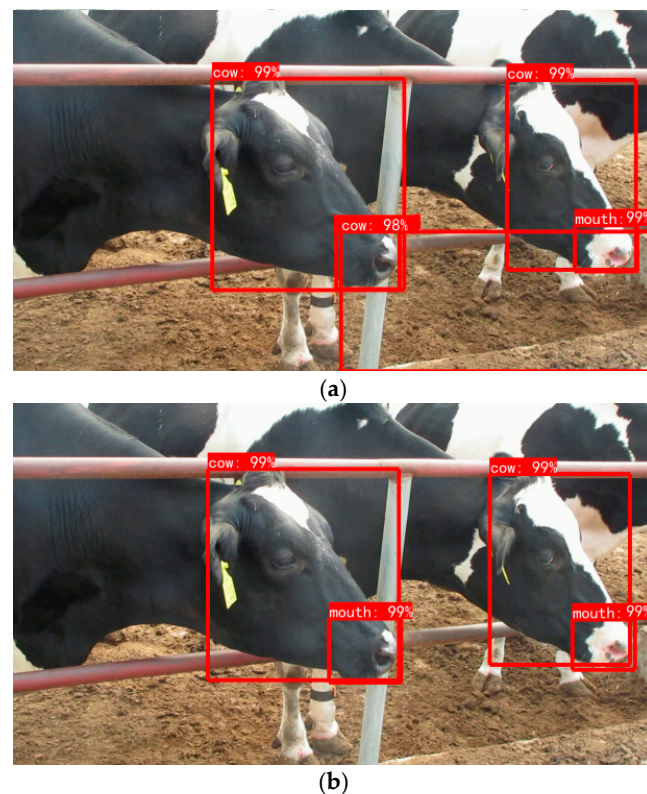


**Figure 11.** Comparison diagram of object detection results. (a) Faster R-CNN detection result of video No. 2; (b) Faster R-CNN detection result of video No. 2; (c) Faster R-CNN detection result of video No. 3; (d) Faster R-CNN detection result of video No. 3; (e) Faster R-CNN detection result of video No. 6; (f) Faster R-CNN detection result of video No. 6.

Figure 11 presents a comparison of the algorithm’s performance before and after improvement in long-distance multi-object ruminant video detection. In Figure 11a, some

of the cow's mouth regions were not successfully detected, resulting in slightly lower detection scores for both the cow's object region and mouth region compared to Figure 11b. However, Figure 11b successfully detects all the cow's mouth objects with improved detection scores. In Figure 11c,d, the detection frame of the cow's mouth region on the right is missing due to railing obstruction. On the other hand, Figure 11e,f show the detection results of the multi-object ruminant video at close range. Notably, Figure 11f demonstrates that the improved algorithm maintains a high detection success rate and detection scores even under suboptimal lighting conditions.

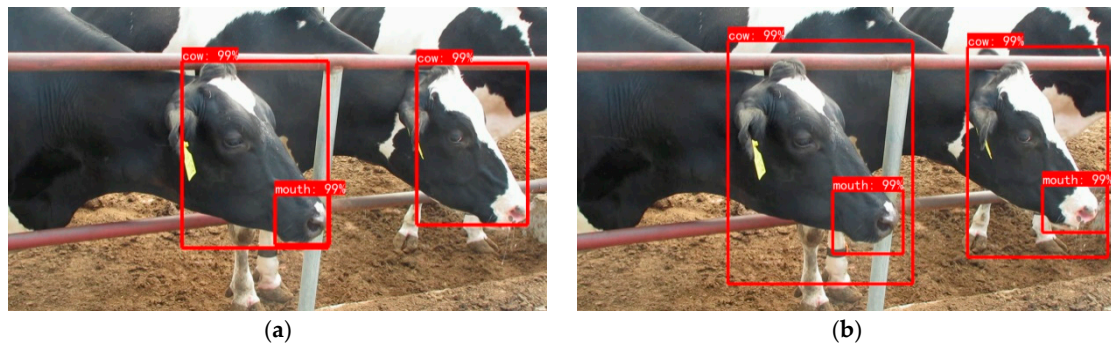
To summarize, the integration of the CBAM attention mechanism into the Faster R-CNN cow regurgitation region detection algorithm has enhanced model performance and detection success rate compared to the original algorithm without any weight surge. Within the dataset used in this study, a detection success rate of 84.70% was achieved for the cow's mouth region. However, in the continuous video frame detection task, incorrect detections and omissions are inevitable. To ensure accuracy and stability in continuous video frame detection, this study combines the optical flow algorithm with the complementary frame algorithm based on interpolation to further correct the detection results. To evaluate the impact of the fused optical flow algorithm, experimental results comparing the original algorithm with the fused FlowNet 2.0 optical flow algorithm are presented in Figure 12.



**Figure 12.** Example of eliminating the wrong detection area. (a) Original target detection result; (b) Corrected result after fusion of optical flow information.

As observed in Figure 12a, there are no cow objects present within the red boxed area in the lower background. Upon traversing the detection box areas, the optical flow algorithm calculates that the optical flow motion information in the lower right detection box area falls significantly below the set threshold  $\mu'$ . Consequently, the erroneously detected detection box is rejected. Figure 12b illustrates the outcome of this rejection, demonstrating the successful elimination of the wrongly detected object frame in the background region. The object detection algorithm, which incorporates optical flow information, effectively addresses the issue of background misdetection in the original algorithm.

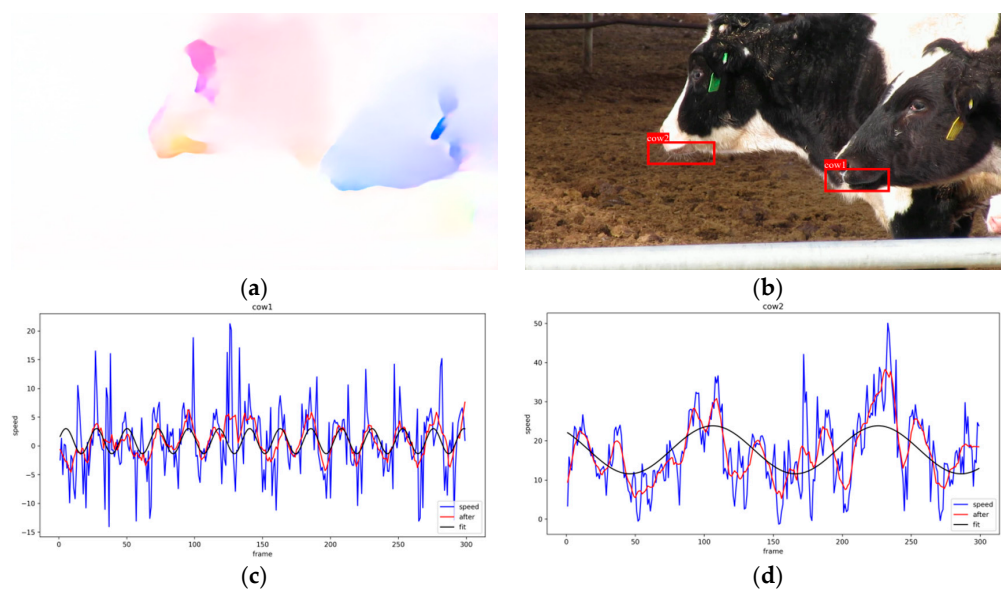
To validate the efficacy of the algorithm's missed frame correction, a comparison was conducted between the original algorithm and the algorithm augmented with the frame-completion algorithm. The results of this comparison are presented in Figure 13.



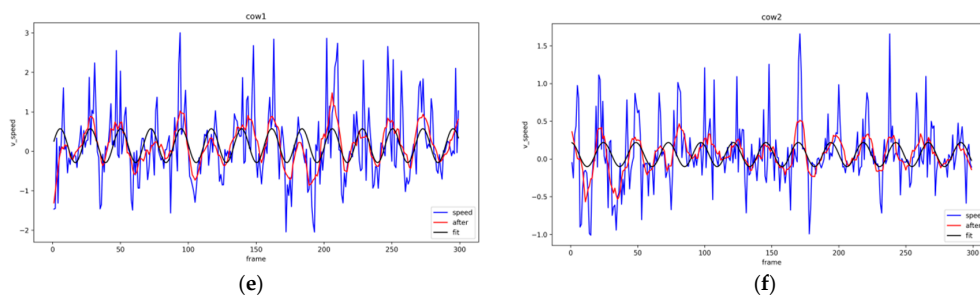
**Figure 13.** Example of correcting missed detection. (a) Original target detection result; (b) Result after adding the complementary frame algorithm.

Figure 13 illustrates the successful supplementation of the missing mouth detection frame of the cow object located on the right side of Figure 13a with a detection reference frame, as demonstrated in Figure 13b. Additionally, the isometric expansion of the detection frame ensures the capture of complete mouth optical flow information. By integrating the interpolation-based complementary frame algorithm, a corresponding detection frame is generated for each frame of the multi-object cow mouth region in the detection result. This ensures the continuity of the detection frame for the mouth region in the video frame sequence, ultimately facilitating the detection of the cow regurgitation region in the multi-object scene.

To validate the robustness and effectiveness of the improved GMFlowNet algorithm, which incorporates Mean-Shift clustering, a comparison was made between the light flow curves generated with the FlowNet 2.0 algorithm and the improved GMFlowNet algorithm on the same device for analyzing ruminant behavior. This comparison is presented in Figure 14.



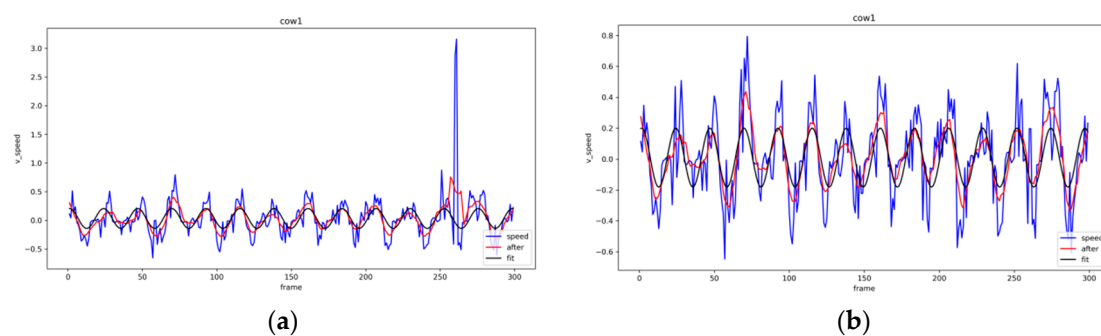
**Figure 14.** Cont.



**Figure 14.** Comparison of optical flow curve with different algorithms. (a) Video optical flow diagram; (b) Multi-object mouth area detection results; (c) FlowNet 2.0 Cow1 optical flow curve; (d) FlowNet 2.0 Cow2 optical flow curve; (e) GMFlowNet Cow1 optical flow curve; (f) GMFlowNet Cow2 optical flow curve.

The optical flow graph utilizes the sequence number of video frames as the horizontal coordinate and the vertical velocity component of the optical flow as the vertical coordinate. Figure 14b demonstrates that the frame for detecting the regurgitation area in multi-object cows is positioned near the object's mouth, resulting in a reduced background area. This ensures the reliability of the extracted optical flow information. Figure 14c,e display the optical flow curves of cow1's object mouth movement, which is marked on the right side of Figure 14b. Cow1, being closer to the filming device, generates accurate and detailed optical flow information. Similarly, Figure 14d,f exhibit the optical flow curves of cow's object mouth movement, identified on the left side of Figure 14b. The improved GMFlowNet algorithm proficiently extracts small-scale optical flow information, yielding curves that demonstrate significant periodicity and prominent peaks and valleys. Consequently, the algorithm ensures both robustness and accuracy. In conclusion, the enhanced GMFlowNet algorithm effectively captures the regurgitation movement information of small-scale multi-object cows, even at longer distances.

To validate the effectiveness of the outlier replacement logic, which incorporates the concept of mean squared deviation, the optical flow curves generated with the GMFlowNet algorithm were compared with those produced with the algorithm augmented with the outlier replacement module on the same device. The results are depicted in Figure 15.



**Figure 15.** Comparison diagram of optical flow curve results. (a) Original optical flow curve; (b) Optical flow curve after outlier replacement.

Figure 15 displays the original optical flow curve in Figure 15a, wherein a significant outlier is observed at frame 265 due to interference caused by the cow's head movement in the original video. Consequently, the curve exhibits insignificant peaks and valleys, rendering it susceptible to errors during period calculation. Conversely, Figure 15b shows the optical flow curve after replacing the outlier, demonstrating improved periodicity and clearer peaks and valleys. The above comparative analysis verifies that replacing outliers can reduce interference during motion curve analysis and ensure computational robustness.

In order to further validate the accuracy of the enhanced GMFlowNet algorithm in identifying the reflux behavior of multi-objective cows and calculating the chewing frequency of cows, a comparative analysis was conducted between the FlowNet 2.0 algorithm and the improved GMFlowNet algorithm. Nine videos featuring multi-object cow regurgitation in diverse scenes were utilized for this purpose. Table 3 provides a comprehensive summary of the comparison results.

**Table 3.** Comparison of analysis results of ruminant behavior in dairy cows.

Video Serial Number	Dairy Cow Number	To Ruminant or Not to Ruminant	Actual Number of Ruminant Chews	FlowNet 2.0 Algorithm Analysis Results		GMFlowNet Algorithm Analysis Results	
				Ruminant Behavior Judgement	Calculated Value of Ruminant Chewing Times	Ruminant Behavior Judgement	Calculated Value of Ruminant Chewing Times
1	Cow1	✓	14	✓	14	✓	14
	Cow2	✓	13	×	-	✓	13
2	Cow1	✓	11	✓	11	✓	11
	Cow2	✓	11	✓	12	✓	11
3	Cow1	✓	11	✓	12	✓	11
	Cow2	×	-	×	-	×	-
4	Cow1	✓	13	✓	13	✓	13
	Cow2	✓	12	×	-	✓	12
	Cow3	×	-	×	-	×	-
5	Cow1	✓	12	×	-	✓	12
	Cow2	✓	9	✓	6	✓	13
	Cow3	×	-	×	-	×	-
6	Cow1	✓	13	×	-	✓	13
	Cow2	✓	13	✓	13	✓	13
	Cow3	×	-	×	-	×	-
7	Cow1	×	-	×	-	×	-
	Cow2	✓	16	✓	16	✓	16
	Cow3	×	-	×	-	×	-
	Cow4	×	-	×	-	×	-
8	Cow1	✓	11	✓	12	✓	10
	Cow2	✓	12	✓	10	✓	12
9	Cow1	✓	14	✓	14	✓	14
	Cow2	×	-	×	-	×	-
Accuracy of regurgitation behavior analysis/%				82.61	93.33	100.00	97.30

As observed in Table 3, the FlowNet2.0 algorithm achieved an accuracy of 82.61% in determining ruminant behavior, whereas the improved GMFlowNet algorithm correctly determined all ruminant behaviors. Regarding the detection of ruminant chewing count, the FlowNet2.0 algorithm attained an accuracy of 93.33%, whereas the improved GMFlowNet algorithm achieved an accuracy of 97.30%, representing an improvement of 3.97 percentage points. It should be noted that in video 5, cow2 ceased ruminating in the second half of the recording, resulting in a higher false detection rate in the curve fitting process. Nonetheless, the improved GMFlowNet algorithm consistently demonstrated high accuracy rates for ruminant behavior determination and ruminant chewing count detection across the experimental dataset. These results validate the stability and robustness of the improved GMFlowNet algorithm.

## 6. Conclusions

The method proposed in this paper is based on image detection technology that can realize automatic monitoring and analysis of cow rumination behavior in large-scale breeding. Traditional monitoring of cow rumination behavior requires manual observation

and recording, which is inefficient and prone to errors. The method based on image detection can be monitored in an automatic way, which greatly improves the efficiency and accuracy of monitoring. At the same time, the method based on image detection can realize non-contact monitoring, that is, there is no need for any intervention on the cow, only the need to install a camera for monitoring. This not only reduces interference to cows but also allows real-time monitoring of a large number of cows, which is suitable for large-scale farming scenarios. In addition, the method based on image detection can also build an intelligent monitoring system, which can detect anomalies in time and make corresponding adjustments by analyzing the data of cows' rumination behavior. For example, when a cow's rumination behavior deviates significantly from the normal range, the system can automatically sound an alarm so that breeders can take timely measures. Therefore, the method proposed in this paper is of great significance for the realization of large-scale breeding and the construction of non-contact intelligent image detection system, and it provides technical support for the intelligent monitoring and analysis of cow rumination behavior in group breeding modes.

The purpose of this study is to explore several problems in the monitoring of rumination behavior, including high labor costs, animal stress, complex farming environments, and the need for non-contact analysis methods in multi-objective scenarios. To overcome these challenges, the study investigates precise detection and movement analysis methods for multi-objective cow regurgitation behavior using the optical flow method and object detection algorithm. An improved Faster R-CNN method is proposed for detecting regurgitation regions in multiple object scenarios, specifically focusing on accurately detecting the cow's regurgitation mouth regions. To enhance the detection accuracy, the CBAM attention mechanism and ResNet-50-FPN network are incorporated to extract features from the cow's mouth. Additionally, the improved FlowNet2.0 optical flow algorithm is integrated into the object detection process to identify and reject false detections caused by the algorithm itself, reducing the false detection rate. The interpolation-based complementary frame algorithm is used to supplement and filter the detected frames of the mouth region, addressing the issue of missed detections and improving the accuracy of cow mouth region detection. Experimental results demonstrate that the average accuracy of the improved Faster R-CNN algorithm is 84.70%, which is 11.80 percentage points higher than the accuracy before the improvements. Furthermore, an improved GMFlowNet optical flow algorithm is proposed for multi-object cow ruminant behavior analysis. This algorithm calculates the multi-object optical flow by utilizing the vertical optical flow component instead of the optical flow velocity, thus eliminating the interference caused by partial head motion on the extraction of multi-object optical flow. To enhance the accuracy and efficiency of smaller-scale optical flow estimation, the Mean-Shift clustering method is integrated to calculate the optical flow information and traverse the video frame sequence to derive the multi-object cow ruminant behavior. In order to eliminate the influence of abnormal values on curve pattern analysis, the mean square difference is computed and combined with the interquartile spacing concept to remove abnormal optical flow data. After filtering and fitting, the calculation of the number of regurgitated chewing cows is achieved. The experimental results show that the improved GMFlowNet algorithm has an accuracy of 97.30% in chewing times, which is 3.97 percentage points higher than the FlowNet 2.0 algorithm, based on accurately identifying the rumination behavior of multi-objective cows. This improved optical flow algorithm successfully addresses the challenges of extracting optical flow from small-scale multi-object cow ruminant behavior, eliminates the interference of abnormal data on cow ruminant curve analysis, and ensures the accuracy and robustness of the algorithm. The proposed method presented in this study is significant for achieving large-scale breeding and developing a non-contact intelligent image detection system. It provides technical support for the intelligent monitoring and analysis of cow regurgitation behavior in group breeding modes.

**Author Contributions:** Conceptualization, R.G.; methodology, R.G.; software, Q.L. (Qihang Liu); validation, Q.L. (Qihang Liu); formal analysis, R.G.; investigation, Q.L. (Qihang Liu); resources, Q.L. (Qifeng Li); data curation, Q.L. (Qifeng Li); writing—original draft preparation, Q.L. (Qihang Liu) and L.Y.; writing—review and editing, Q.L. (Qihang Liu), L.Y. and J.J.; visualization, Q.B.; supervision, J.J. and Q.L. (Qifeng Li); project administration, K.Z.; funding acquisition, R.G. and Q.L. (Qifeng Li). All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by Beijing Academy of Agricultural and Forestry Sciences Science and Technology Innovation Capacity Construction Project (KJ CX20220404, KJ CX20230204), and National Natural Science Foundation of China (Grant No. 32002227).

**Data Availability Statement:** As the data used in this manuscript will also be used for other technical research, it is currently not publicly available.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Han, L.; Liu, C.Q. Development trends, challenges and policy recommendations of China's dairy economy. *China J. Anim. Husb.* **2019**, *55*, 151–156. [[CrossRef](#)]
- Wei, J.X. Analysis of dairy farming patterns and their efficiency in China in the new era. *China Herbiv. Sci.* **2018**, *38*, 49–51.
- He, D.J.; Liu, D.; Zhao, K.X. Research progress on intelligent animal information perception and behavior detection in precision animal husbandry. *J. Agric. Mach.* **2016**, *47*, 231–244.
- Liu, Z.C.; Zhai, T.C.; He, D.J. Current status and progress of research on individual information monitoring of dairy cows in precision farming. *Heilongjiang Anim. Husb. Vet. Med.* **2019**, *13*, 30–33+38. [[CrossRef](#)]
- Wang, S.; Hu, F.M.; Diao, Q.Y.; Tu, Y. Research progress on the regulation mechanism of ruminant behavior in dairy cows. *J. Anim. Nutr.* **2021**, *33*, 1869–1879.
- He, D.Q.; Tao, J.Z. Differences in rumen function and related molecular biological mechanisms in residual feed intake of ruminants. *J. Anim. Nutr.* **2021**, *33*, 3125–3131.
- Wang, Y.; Lu, N.; Wang, Y.; Ma, Y. Research progress on intelligent monitoring method of dairy cattle ruminant behavior and its application. *China Feed* **2021**, *7*, 3–6. [[CrossRef](#)]
- Bauermann, F.V.; Joshi, L.R.; Mohr, K.A.; Kutish, G.F.; Meier, P.; Chase, C.; Christopher-Hennings, J.; Diel, D.G. A novel bovine papillomavirus type in the genus *Dyokappapapillomavirus*. *Arch. Virol.* **2017**, *162*, 3225–3228. [[CrossRef](#)] [[PubMed](#)]
- Feng, Y.; Gao, Z.T.; Zheng, W.B.; Yang, Z.T.; Dong, Q. Application of machine learning for clinical disease prediction in dairy cattle. *Adv. Anim. Med.* **2021**, *42*, 115–119. [[CrossRef](#)]
- Kruse, K.M.; Combs, D.K.; Beauchemin, K.A. Effects of forage particle size and grain fermentability in midlactation cows. I. Milk production and diet digestibility. *J. Dairy Sci.* **2002**, *85*, 1936–1946. [[CrossRef](#)]
- Shao, D.F. *A Study on the Correlation between the Change Pattern of Ruminant Behavior and Its Influencing Factors in Dairy Cows*; Jilin University: Changchun, China, 2015.
- Liang, G.F.; Li, Y.; Bai, D.N.; Sun, X.T.; Lei, X.J.; Cai, C.J.; Yao, J.H.; Cao, Y.C. Advances in ruminant starch nutrition. *J. Livest. Ecol.* **2021**, *42*, 7–12.
- Li, Y.T.; Yan, R.; Hu, Z.Y.; Lin, X.Y.; Wang, Z.H. Effect of particle size of total mixed diets on feeding, chewing activity and production performance of dairy cows. *J. Anim. Nutr.* **2017**, *29*, 298–308.
- Zhang, S.L.; Wang, R.Y.; Shi, G.F.; Cheng, S.L.; Xin, Y.P. Effect of roughage length on chewing behavior and production performance of dairy cows. *J. Anim. Husb. Vet. Med.* **2017**, *36*, 29–31.
- Zhu, D.P.; Liu, C.; Hu, Z.B.; Lin, M.Q.; Wu, F.Q.; Zhou, Q.X.; Liu, Z.C.; Zhang, W.; Yang, H.J.; Mo, F. Comparative analysis of feeding, ruminant and prolific behavior and apparent digestibility of dietary fibrous matter in lactating dairy cows. *Livest. Feed Sci.* **2019**, *40*, 30–34. [[CrossRef](#)]
- Stone, A.E.; Jones, B.W.; Becker, C.A.; Bewley, J. Influence of breed, milk yield, and temperature-humidity index on dairy cow lying time, neck activity, reticulorumen temperature, and rumination behavior. *J. Dairy Sci.* **2017**, *100*, 2395–2403. [[CrossRef](#)] [[PubMed](#)]
- Kaufman, E.; Asselstine, V.; Leblanc, S.; Duffield, T.; DeVries, T. Association of rumination time and health status with milk yield and composition in early-lactation dairy cows. *J. Dairy Sci.* **2018**, *101*, 462–471. [[CrossRef](#)] [[PubMed](#)]
- Antanitis, R.; Zilaitis, V.; Juozaitiene, V.; Noreika, A.; Rutkauskas, A. Evaluation of rumination time, subsequent yield, and milk trait changes dependent on the period of lactation and reproductive status of dairy cows. *Pol. J. Vet. Sci.* **2018**, *21*, 567–572. [[CrossRef](#)]
- Xiong, A.R.; Wang, H.; Wu, F.X. Analysis of the correlation between milk production and rumination time, activity and somatic cell score in periparturient cows. *J. Anim. Nutr.* **2020**, *32*, 5293–5301.
- Almeida, P.E.; Weber, P.S.D.; Burton, J.L.; Zanella, A. Depressed DHEA and increased sickness response behaviors in lame dairy cows with inflammatory foot lesions. *Domest. Anim. Endocrinol.* **2008**, *34*, 89–99. [[CrossRef](#)] [[PubMed](#)]

21. Si, J.F.; Liu, Y.P.; Dong, G.F.; Zhu, K.C.; Huang, H.T. Predicting the risk of peripartum ketosis in dairy cows using prepartum ruminal rumen regurgitation time. *J. China Agric. Univ.* **2019**, *24*, 78–87.
22. Bao, Y.T.; Chen, X.F.; Zhang, L. Exploration of differential diagnosis of ruminant disorders in dairy cattle. *Agric. Dev. Equip.* **2016**, *164*, 159.
23. Wang, H. Differential diagnosis of ruminant disorders in dairy cattle. *Anim. Husb. Vet. Sci. Technol. Inf.* **2015**, *5*, 74.
24. Chang, S.; Tang, S.Y. Diagnosis and prevention of foregut flaccidity and related diseases in ruminant livestock. *Anim. Husb. Vet. Sci. Technol. Inf.* **2013**, *5*, 50–51.
25. Chacher, B. *Evaluation of n-Carbamoylglutamate as Arginine Enhancer and Its Effect on Rumen Fermentation, Lactation Performance and Nitrogen Utilization in High Yielding Lactating Dairy Cows*; Zhejiang University: Hangzhou, China, 2013.
26. Wang, L.W.; Xie, Q.J.; Liu, H.G.; Yan, L.; Xu, Z.D. Research on a wearable monitoring device for dairy cows' rumination behavior based on multi-source information perception. *Heilongjiang Anim. Husb. Vet. Med.* **2019**, *47–51*, 164–165. [[CrossRef](#)]
27. Zhang, X.S.; Zhang, M.J.; Wang, L.; Luo, H.; Li, J. Research status and development analysis of wearable information monitoring technology for animal husbandry. *J. Agric. Mach.* **2019**, *50*, 1–14.
28. Qin, L.F.; Zhang, X.Q.; Dong, X.M.; Yue, S. Multi-feature fusion correlation filter-based object extraction for sport cows. *J. Agric. Mach.* **2021**, *52*, 244–252.
29. Li, D.; Chen, Y.F.; Li, X.J.; Pu, D. Research progress on the application of computer vision technology in pig behavior recognition. *China Agric. Sci. Technol. Her.* **2019**, *21*, 59–69. [[CrossRef](#)]
30. Song, H.B.; Niu, M.T.; Ji, C.H.; Li, Z.N.; Zhu, Q.M. Multi-objective monitoring of cow ruminant behavior based on video analysis. *J. Agric. Eng.* **2018**, *34*, 211–218.
31. You, X.H.; Ma, Q.; Guo, H.; Wang, Q. Analysis and prospect of dairy cow identification and behavior sensing technology. *Comput. Appl.* **2021**, *41*, 216–224.
32. Burfeind, O.; Schirmann, K.; von Keyserlingk, M.; Veira, D.; Weary, D.; Heuwieser, W. Technical note: Evaluation of a system for monitoring rumination in heifers and calves. *J. Dairy Sci.* **2011**, *94*, 426–430. [[CrossRef](#)] [[PubMed](#)]
33. Yao, Y. *Research on ANT-Based Dairy Cattle Rumination Information Collection System*; Donghua University: Shanghai, China, 2015.
34. Braun, U.; Trosch, L.; Nydegger, F.; Hassig, M. Evaluation of eating and rumination behaviour in cows using a noseband pressure sensor. *BMC Vet. Res.* **2013**, *9*, 164. [[CrossRef](#)] [[PubMed](#)]
35. Zehner, N.; Umstaetter, C.; Niederhauser, J.J.; Schick, M. System specification and validation of a noseband pressure sensor for measurement of ruminating and eating behavior in stable-fed cows. *Comput. Electron. Agric.* **2017**, *136*, 31–41. [[CrossRef](#)]
36. Chen, Y.; He, D.; Fu, Y.; Song, H. Intelligent monitoring method of cow ruminant behavior based on video analysis technology. *Int. J. Agric. Biol. Eng.* **2017**, *10*, 194–202. [[CrossRef](#)]
37. Reiter, S.; Sattlecker, G.; Lidauer, L.; Kickinger, F.; Öhlschuster, M.; Auer, W.; Schweinzer, V.; Klein-Jöbstl, D.; Drillich, M.; Iwersen, M. Evaluation of an ear-tag-based accelerometer for monitoring rumination in dairy cows. *J. Dairy Sci.* **2018**, *101*, 3398–3411. [[CrossRef](#)] [[PubMed](#)]
38. Mao, Y.R.; Niu, T.; Wang, P.; Song, H.B.; He, D.J. Multi-objective cow mouth tracking and regurgitation monitoring using Kalman filtering and Hungarian algorithm. *J. Agric. Eng.* **2021**, *37*, 192–201.
39. Zhao, K.X.; He, D.J.; Wang, E.Z. Video-based analysis of cow breathing frequency and abnormality detection. *J. Agric. Mach.* **2014**, *45*, 258–263.
40. Song, H.B.; Wu, D.H.; Yin, X.Q.; Jiang, B.; He, D.J. Respiratory behavior detection of dairy cows based on Lucas-Kanade sparse optical flow algorithm. *J. Agric. Eng.* **2019**, *35*, 211324.
41. Li, T. *Research on Multi-Objective Cow Ruminant Behavior Monitoring Method Based on Video Analysis*; Northwest Agriculture and Forestry University of Science and Technology: Yangling, China, 2019.
42. Ji, J.T.; Liu, Q.H.; Gao, R.H.; Li, Q.F.; Zhao, K.X.; Bai, Q. Analysis method of cow rumination behavior based on improved flownet 2.0 optical flow algorithm. *J. Agric. Mach.* **2023**, *54*, 235–242.
43. Fu, B.; Huang, Z.B.; Quan, Y.; Feng, W.L. Secondary correction method of deformable instrument image recognition based on fast regional Convolutional neural network. *Sci. Technol. Eng.* **2023**, *23*, 9122–9129.
44. Cui, Y.B.; Tang, R.D.; Xing, D.J.; Wang, W.; Li, S.S. Visual optical flow computing technology and its application. *J. Electron. Inf. Technol.* **2023**, *45*, 2710–2721.
45. Sheng, W.; Yu, X.; Lin, J.; Chen, X. Faster RCNN Target Detection Algorithm Integrating CBAM and FPN. *Comput. Syst. Sci. Eng.* **2023**, *47*, 1549–1569. [[CrossRef](#)]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.