Biosystems Engineering 246 (2024) 54–66

Contents lists available at ScienceDirect

Biosystems Engineering

journal homepage: www.elsevier.com/locate/issn/15375110

Research Paper

IATEFF-YOLO: Focus on cow mounting detection during nighttime

De Li [a], Baisheng Dai [a,*], Yanxing Li [a], Peng Song [a], Xin Dai [a], Yongqiang He [c], Huixin Liu [a], Yang Li [b], Weizheng Shen [a,**]

[a] College of Electrical Engineering and Information, Northeast Agricultural University, Harbin, 150030, China
[b] College of Animal Sciences and Technology, Northeast Agriculture University, Harbin, 150030, China
[c] Inner Mongolia Mengniu Dairy (Group) CO., LTD, Shengle Economic District Helingeer, Hohhot, Inner Mongolia, 011517, China

ARTICLE INFO

Keywords:
Cow oestrus
Cows mounting behaviour detection
Computer vision
IATEFF-YOLO

ABSTRACT

Mounting behaviour is an important characteristic of cows during oestrus. Real-time and accurate detection of cow mounting behaviour can shorten the calving-to-conception period and increase the economic benefits for dairy farms. Cow mounting behaviour occurs more often at night, and drastic scale changes in surveillance images caused by different distances between cows and camera, influence the detection of cow mounting. Existing methods are unable to address these challenges effectively. To address these challenges, this study collected 9392 images of Holstein cow mounting behaviour under intensive farming conditions using cameras and proposed an IATEFF-YOLO that is more suitable for cow mounting behaviour detection at nighttime and drastic scale changes in surveillance images caused by different distances between cows and camera. IATEFF-YOLO comprises an Illumination Adaptive Transformer (IAT) and an efficient feature fusion detector. The IAT enhances low-light images at night to enrich the cow mounting features, facilitating the subsequent detection of mounting behaviour. The efficient feature fusion detector, EFF-YOLO, enhances the feature fusion capability and further enable the detector to adapt to drastic scale changes in surveillance images caused by different distances between cows and camera. IATEFF-YOLO achieved a mean Average Precision of 99.3% and a detection speed of 102.0 f/s on test set. Compared with existing methods, IATEFF-YOLO achieved higher detection accuracy and faster detection speed during nighttime and drastic scale changes in surveillance images caused by different distances between cows and camera. IATEFF-YOLO can assist ranch breeders in achieving round-the-clock monitoring of cow oestrus, thereby enhancing oestrus detection efficiency.

Nomenclature

Abbreviations	
CSPC	Cross Stage Partial Connection
DETR	DEtection TRansformer
FPN	Feature Pyramid Network
IAT	Illumination-Adaptive-Transformer
mAP	mean Average Precision
PEM	Pixel-wise Enhancement Module
Pr	Precision
Re	Recall
SPP	Spatial Pyramid Pooling
SSD	Single Shot MultiBox Detector
YOLO	You Only Look Once

1. Introduction

Real-time and accurate detection of oestrus in dairy cows is crucial for improving conception rates, shortening the calving-to-conception interval, and enhancing the economic benefits of dairy farms (Van Vliet & Van Eerdenburg, 1996). The close relationship between cow oestrus and mounting behaviour, and timely detection of cow mounting behaviour is important for monitoring cow oestrus (Yusheng et al., 2020). Mounting behaviour is an important manifestation of the dairy cow's oestrus (Mwaanga & Janowski, 2000). However, traditional manual detection of cow mounting behaviour is time-consuming and laborious, failing to meet the needs of modern breeding. Therefore, the study of automatic and real-time detection of cow mounting behaviour is of great significance.

With the rapid development of sensor and computer vision

* Corresponding author.
** Corresponding author.
E-mail addresses: bsdai@neau.edu.cn (B. Dai), wzshen@neau.edu.cn (W. Shen).

https://doi.org/10.1016/j.biosystemseng.2024.07.017
Received 15 February 2024; Received in revised form 20 July 2024; Accepted 21 July 2024
Available online 27 July 2024

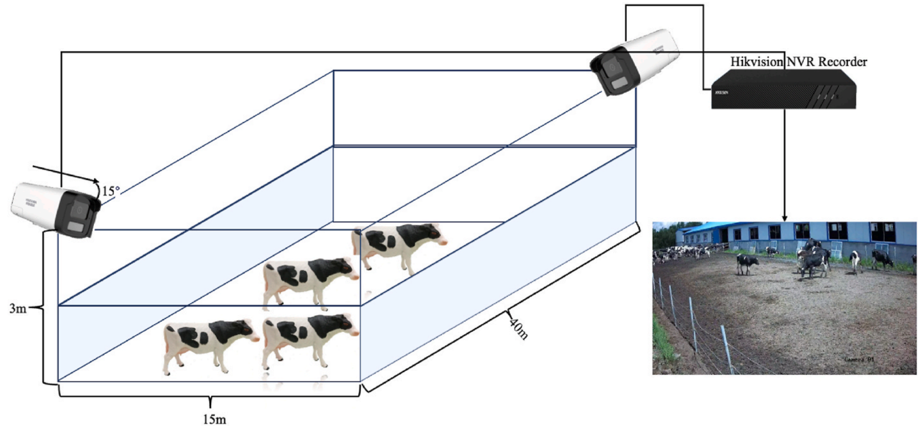

Fig. 1. Schematic of data collection system.

technology, detection methods of oestrus behaviour in cows can be divided into contact and non-contact methods (Wang, Bai, et al., 2022). In the method of contact detection of oestrus behaviour in dairy cows, it is usual to wear collars, foot ring and in-ear tags on cows to monitor the activity information or temperature of cows so as to infer the oestrus of cows (Higaki et al., 2021; MacKay et al., 2012; Perez et al., 2023; Wang, Liao, et al., 2020). The contact detection of oestrus behaviour in dairy cows is likely to cause a stress response to the cows, and the sensor worn by the cows will also be damaged due to the activities of the cows, resulting in the failure to detect the oestrus of cows (Wang et al., 2022, 2023). Additionally, deploying many sensors increases the operating costs of dairy farms. For this reason, in order to reduce the stress response of dairy cows, some researchers have proposed using computer vision technology for non-contact oestrus detection in dairy cows (Lodkaew et al., 2023; Pasupa & Lodkaew, 2019; Sumi et al., 2018). The mounting behaviour of cows pertains to the act where one heifer ascends upon another from the rear, with the latter remaining stationary or struggling to maintain equilibrium beneath the weight, accompanied by pelvic thrusts from the mounting heifer (Esslemont et al., 1980). There exists a profound link between dairy cow oestrus and mounting behaviour, making the timely detection of mounting behaviour in dairy cows a pivotal consideration for identifying oestrus and enhancing farming profitability (Roelofs et al., 2005; Roelofs et al., 2010; Van Vliet & Van Eerdenburg, 1996).

Non-contact oestrus detection mainly depends on detecting mounting behaviour (Chung et al., 2015). Tsai and Huang (2014) proposed determining the movement area of cows and using foreground segmentation to separate moving cows from the background, detecting mounting behaviour based on changes in moving cow length. Guo et al. (2019) proposed a method for detecting cow mounting behaviour using the geometric and optical flow characteristics by removing irrelevant backgrounds, extracting features, and classifying mounting cows with support vector machines. Chae and Cho (2021) improved YOLOV3 (Redmon & Farhadi, 2018) object detection network by adding an additional layer and a new activation function to improve the ability of YOLOV3 to detect cow mounting behaviour. Wang, Bai, et al. (2022) proposed a lightweight cow oestrus behaviour detection model combining YOLOV5n (Jocher et al., 2021) and a channel pruning algorithm to improve detection accuracy.

These non-contact methods have made good progress, but there are still some challenges. A significant proportion (62.1%) of cow mounting behaviour occurs at night (Mwaanga & Janowski, 2000). The nighttime is a high-occurrence period for cows' mounting behaviour which insufficient lighting, the video images captured by surveillance cameras during this time are often low-light. This low-light condition significantly reduces the clarity of surveillance videos images, making it difficult to identify key details of cows' mounting behaviour. Additionally, as cows move, the distance between them and surveillance cameras constantly changes, causing drastic scale variations in the video images, complicating accurate detection.

To address these challenges, this study proposes the IATEFF-YOLO model for efficient cow mounting behaviour detection. IATEFF-YOLO includes an Illumination Adaptive Transformer (Cui et al., 2022) and an efficient feature fusion detector (EFF-YOLO). The IAT enhances low-light images to facilitate nighttime detection. EFF-YOLO, based on the YOLOX-S framework (Ge et al., 2021), incorporates an SPPCSPC block for improved multiscale feature fusion and a Selective Kernel Attention Decoupled Head, based on Selective Kernel Attention (Li et al., 2019), to adapt to drastic scale changes in surveillance images caused by different distances between cows and camera. These improvements enable IATEFF-YOLO to detect cow mounting behaviour at nighttime and drastic scale changes in surveillance images caused by different distances between cows and camera, meeting the real-time detection requirements.

The main contributions of this paper are:

- Addressed the challenge of detecting cows mounting behaviour in nighttime low-light conditions by integrating IAT.
- Tackled the challenge of detecting cows mounting behaviour in multi-scale environments by integrating EFF-YOLO.
- Provided a dataset support for precision livestock research and animal behaviour studies.

2. Materials and methods

2.1. Data sources

Since cow mounting behaviour mainly occurs in outdoor pasture exercise areas (Wang, Bai, et al., 2022), the design of the data acquisition device in this study was centred on the characteristics of the outdoor exercise area. Through consultation with relevant experts from Shengkang Ranch in Daqing City, Heilongjiang Province, this study found that the outdoor exercise area for dairy cows in the ranch is 40.0 m long and 15.0 m wide. There are three independent outdoor exercise areas on the ranch, and each outdoor exercise area has about 100 cows per herds. Due to the complexity of the dairy cow outdoor exercise area environment, this study chose to install two surveillance cameras (DS-2CD3T46WDV3-I3, Hikvision, China) with a focal length of 6 mm, a 4 million pixels, resolution of 2560 × 1440 pixels, and frame rate of 25 fps at the diagonal position of the dairy cow outdoor exercise area. This study adopts a top-down perspective to capture a more complete view of the video monitoring of the dairy cow outdoor exercise area. The surveillance cameras are connected to a Hikvision Network Video Recorder (NVR, DS-7804 N–K1, Hikvision, China) via a network cable, forming a

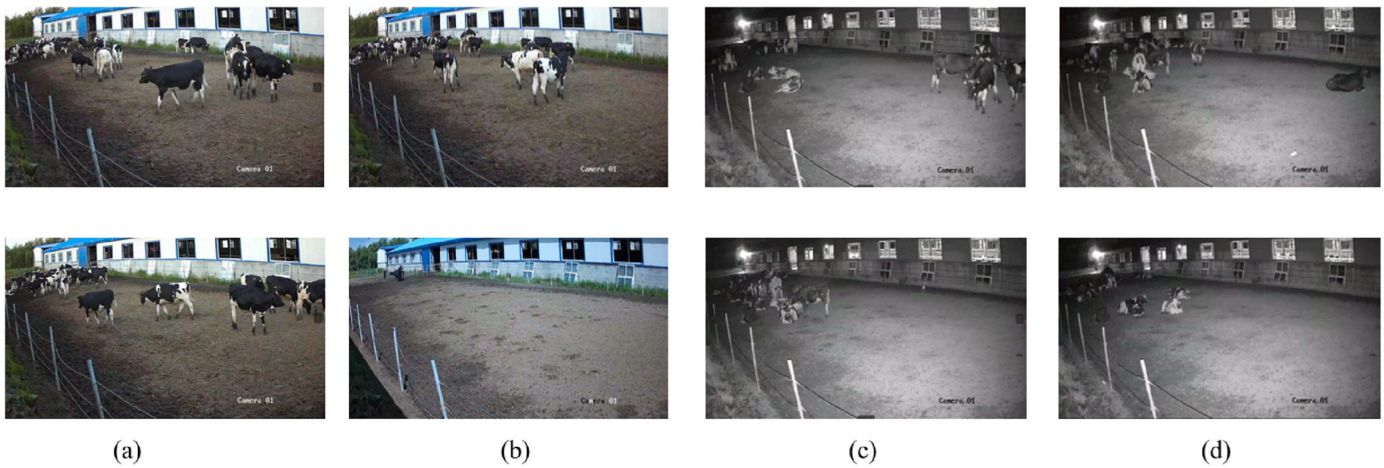

Fig. 2. Examples of datasets are shown. (a) and (b) are images of cow mounting behaviour in a daytime environment. (c) and (d) are images of cow mounting behaviour in a nighttime environment.

complete data collection scheme, as shown in Fig. 1.

This study collected data from the Heifers Holstein (Age of 6–12 months old) cow outdoor exercise area of the Shengkang Animal Husbandry in Lindian County, Daqing City, Heilongjiang Province, in July 2022 and August 2023. Approximately 30 days of surveillance videos were filmed, and approximately a total of 15 TB of video data were collected. Surveillance cameras were installed at the diagonal corners of the outdoor exercise area, positioned 3.0 m above the ground, and tilted downwards at a 15.0° angle, as shown in Fig. 1.

In the Holstein Heifers cows outdoor exercise area where cameras were deployed, there were a total of 107 head of cows that were raised with freedom access to water and feed, vaccinated regularly, and their manure is cleaned periodically. During the data collection process, all cameras were utilised to provide round-the-clock surveillance of dairy cows in their farming feeding environments.

2.2. Data pre-processing

After preprocessing the collected video data, video decomposing frame technology was adopted, taking one frame every five frames to obtain the video frame image, and the video images were normalized, with the image size adjusted to 1920 × 1080 pixels. Cow mounting images were extracted from 360 videos (each video is 1 h long), and a labelling image annotation tool (https://github.com/tzutalin/labelImg) was used to manually mark the cow mounting behaviour area. In this study, 8550 images of cow mounting behaviour to train and evaluate model performance. In addition, to better evaluate the detection ability in the nighttime environments, this paper collected an additional 342 images of cow mounting behaviour in the nighttime environments. Some cow mounting data images are shown in Fig. 2.

2.3. Dataset partition

In this study, 8550 cow mounting images were labelled for the first time and randomly divided into Training set, Validation set, and Test set at a ratio of 6:2:2, resulting in 5130 training examples, 1710 validation examples, and 1710 test examples. In Test, there are nighttime low-light cows mounting behaviour images account for 40.0%. To better test the performance of the model at night, this study additionally constructs a test set called Test-night that removed 1052 non-nocturnal cow mounting images from the Test, and added 342 additional nocturnal cow mounting images. In Test-night, there were 1000 test examples, and 100.0% of the images in the nocturnal environment were cow mounting images. Finally, to better evaluate the robustness of the model in practical applications, a Test-robust dataset containing 500 images was constructed. This test set was collected at FuRui pasture in Harbin, Heilongjiang Province, where the camera's tilt angle was adjusted to 20° based on a previous scheme, and images of Ayrshire cows mounted on the Internet were also collected. Additionally, this study captured relevant images of mounting behaviours using mobile phones. This paper has collected and labelled a total of 9392 images of cows mounting. This paper will use visual method to evaluate the detection performance of the model in drastic scale changes in surveillance images caused by different distances between cows and camera. The cow mounting behaviour detection dataset is shared with the precision livestock farming research community at https://github.com/IPCLab-NEAU/Cows-Mounting-Behavior-Detection.

2.4. Data augmentation

Data augmentation is an efficient method for enhancing model generalisation and improving model detection accuracy (Zoph et al., 2020). To adapt the model to the detection of cow mounting behaviour

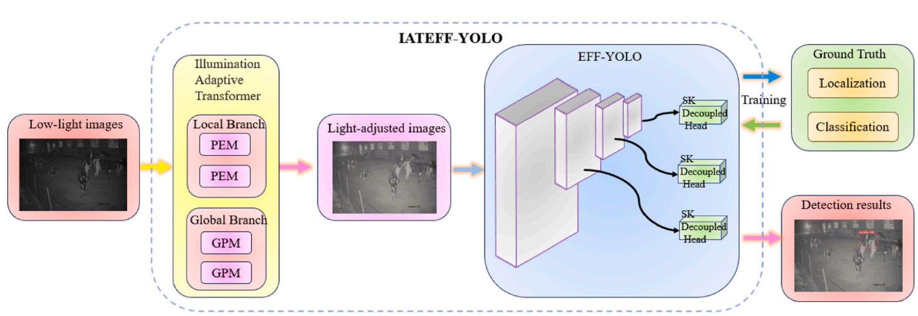

Fig. 3. Structure of the proposed IATEFF-YOLO model.

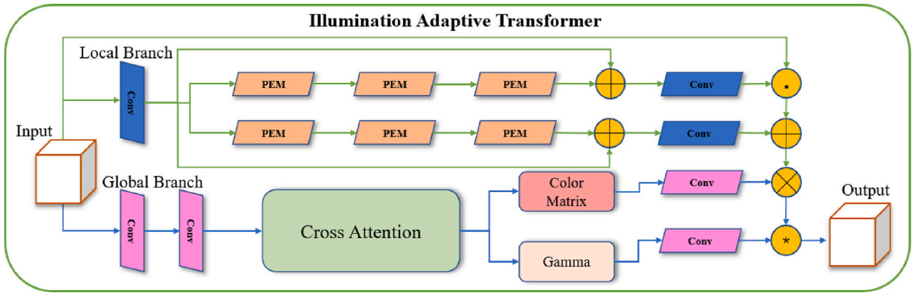

Fig. 4. Structure of illumination adaptive transformer.

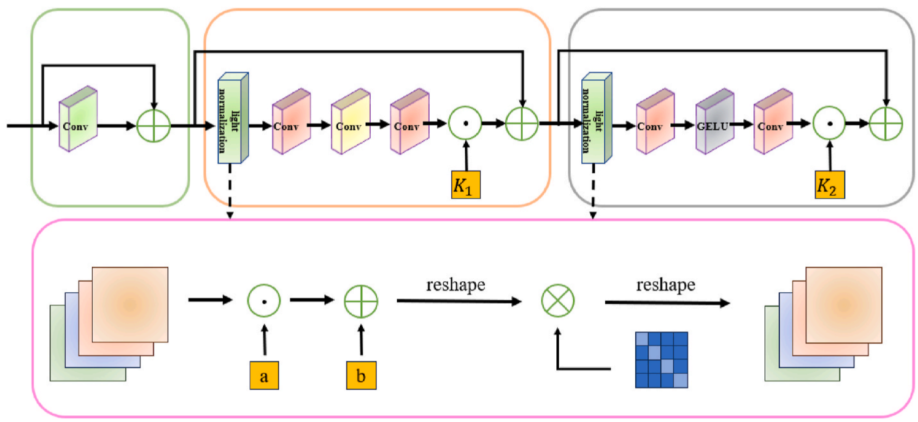

Fig. 5. Structure of PEM module.

in an environment with drastic scale changes in surveillance images caused by different distances between cows and camera and nighttime, Mosaic (Bochkovskiy et al., 2020) data enhancement was used in the cow mounting behaviour training set when model was trained. The data enhancement method randomly selects four images of cow mounting on a splice and obtains an image twice the size of the input image. Subsequently, through random cropping, rotation, translation, scaling, etc., the size of the enhanced image is changed back to an image of the same size as the input image, and finally the mosaic enhancement is realised. The mini-batch uses a portion of the training samples to update the model's parameters. Context features can provide additional information about the target and help the model to identify and locate the target more accurately. Mosaic data augmentation aligns with the mini-batch operation of selecting a smaller subset of data, which can also lead to the incorporation of rich contextual features. Mosaic data augmentation method can reduce the demand for mini-batch and effectively improve the performance of the model.

2.5. IATEFF-YOLO

The IATEFF-YOLO consists of an IAT network (Cui et al., 2022) and an EFF-YOLO detector. The overall structure of IATEFF-YOLO is shown in Fig. 3. First, low-light cow mounting images are enhanced by IAT network, thus enriching the mounting feature extraction to facilitate the subsequent detection of mounting cows. Second, light-adjusted cow mounting images are input to EFF-YOLO to detect cow mounting behaviour.

2.5.1. Illumination Adaptive Transformer

Owing to the low-light conditions at nighttime, the images of dairy cows exhibiting mounting behaviour suffer from decreased image quality. Under such low-light scenarios, it is challenging to extract key features related to the mounting behaviour of cows. These issues can reduce the accuracy of models for detecting mounting behaviours in dairy cows. To meet this challenge, an IAT capable of accurately and real-time enhancing low-light images has been selected to enhance nighttime low-light images of dairy cows mounting. The structure of IAT, which consisting mainly of Local Branch and Global Branch, is shown in Fig. 4. The primary role of Local Branch is to preserve the detailed information in the cow mounting image. The Global Branch enhances the visual effects of the image by adjusting the Color Matrix transform and Gamma correction. After passing through the two branches, the low-light images finally become a light-adjusted images which are beneficial for extracting cow mounting features.

2.5.1.1. Local branch. The Local branch is a lightweight structure that can preserve detailed information about the cow mounting image. First, cow mounting images are expanded in the channel dimension by convolution, and passed to two independent pathways stacked by a Pixel-wise Enhancement Module (PEM). The role of the PEM is to enhance local image details. The structure of the PEM module is shown in Fig. 5. Cow mounting images first encode the positional information using a 3 × 3 depth-wise convolution. Subsequently, the image pixel matrix obtained after encoding is enhanced pixel-wise using two sets of convolutions to enhance the local details.

In the PEM, there is a new normalization called light normalization. Light normalization learns the scales a and bias b through two learnable parameters. In the local branch, three PEM modules were stacked, and the output mounting cow features were concatenated with the input mounting cow features by element-wise addition.

2.5.1.2. Global Branch. The key to enhancing low-light mounting images at night is to adjust the image color and exposure to the obtained image details. In the Global Branch (as shown in Fig. 4), cow mounting images use two convolutions to code the mounting features. Then, the encoded features are input to the Cross Attention block to obtain the parameters for the Color Matrix and Gamma blocks. The Color Matrix

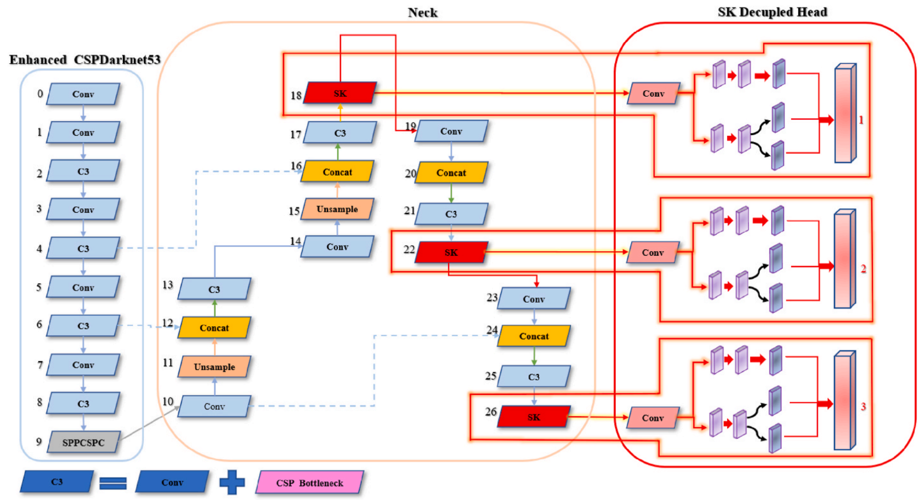

Fig. 6. Structure of the proposed EFF-YOLO.

uses the WB algorithm to estimate the gain of each channel in order to maintain the color constancy of the image. Gamma uses a gamma correction function to correct the low-light problem caused by underexposure at nighttime in cow mounting images. Finally, a light-adjusted mounting image was obtained by performing a multiplication operation, together with the detailed image features obtained by the Local Branch.

2.5.2. EFF-YOLO

When dairy cows are moving, the distance from the surveillance camera fluctuates continuously, causing capture drastic scale changes in the images of cow's mounting behaviour. This leads to significant variations in the mounting features of dairy cows, making it difficult to extract key mounting features accurately. Additionally, the small-scale mounting features of dairy cows can easily be overshadowed by irrelevant background information, causing the model to fail to detect the small-scale mounting behaviour. To address those difficulties, this study proposed an efficient feature fusion detector called EFF-YOLO, which is a novel cow mounting behaviour detection algorithm based on the YOLOX (Ge et al., 2021) model. It inherits many advantages of YOLOX, such as being anchor-free, using a Decouple Head and leading label allocation strategy, SimOTA. EFF-YOLO uses the Enhanced CSPDarkNet53 as the backbone, PANet (Liu et al., 2018) as the Neck, and a Selective Kernel Attention Decoupled Head as the detection head. The structure of EFF-YOLO is shown in Fig. 6, Enhanced CSPDarkNet53 is designed with SPPCSPC block to enhance the multiscale cow mounting feature extraction capability of the backbone module. PANet-Neck is based on the Feature Pyramid Network (FPN) and adds a new bottom-up path to further enhance low-level features, thereby improving the feature diversity obtained by the entire detection network. For a detailed structural introduction, refer to Liu et al. (2018). Subsequently, Selective Kernel Attention Decoupled Head is proposed to further improved the integration of cows mounting key features under multiscale features to achieve accurate detection of cows mounting behaviour.

2.5.2.1. Enhanced CSPDarkNet53. To achieve good feature extraction capabilities for backbone networks in drastic scale changes in surveillance images caused by different distances between cows and camera, SPPCSPC (Wang et al., 2023) was introduced into CSPDarkNet53 (Bochkovskiy et al., 2020) to form an Enhanced CSPDarkNet53. SPPCSPC is a new feature fusion module that combines Spatial Pyramid Pooling (He et al., 2015) and Cross Stage Partial Connection (CSPC) techniques (Wang, Liao, et al., 2020). Spatial Pyramid Pooling (SPP) is a strategy for extracting features at different scales to efficiently handle scale transformation problems and extract multi-scale features. It obtains feature maps of different scales through convolution, then performs MaxPool operations on the different scale feature maps, and finally performs Concat operations on the different scale features. This strategy efficiently handles scale changes problems and extract multi-scale features to detect cow mounting. CSPC enhances the network's ability to learn key cow mounting features by dividing the input cow mounting features into two parts, applying a gradient flow truncation strategy to both branches to avoid gradient information reuse, and integrating cross-layer feature scales to provide a network with rich feature information (Wang, Liao, et al., 2020). The structure of SPPCSPC is shown in Fig. 7.

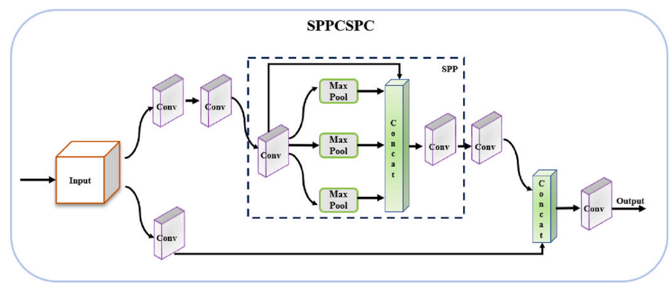

Fig. 7. Structure of SPPCSPC block.

The cow mounting feature map (Input) was obtained through a previous convolution. The cow mounting feature map is divided into two pathways after entering the SPPCSPC block. In the upper branch, the feature map first undergoes convolution and then enter three MaxPool layers with sizes of 5×5, 9×9, and 13×13 in turn. Whenever a new feature map is obtained through a MaxPool layer, in addition to continuing to input to the next MaxPool layer, it will also be directly fed to Concat, and then the Concat operation will be performed together with the feature map that has not gone through MaxPool. In the lower branch, the feature map is first convolved and directly conforms to the feature map obtained by the upper branch for the final output of the SPPCSPC block.

2.5.2.2. Selective Kernel Attention Decoupled Head. To further improve the detection capability of the model in drastic scale changes in surveillance images caused by different distances between cows and camera, a Selective Kernel Attention Decoupled Head was integrated into the EFF-YOLO. The Selective Kernel Attention Decoupled Head is a new detection head based on the original YOLOX Decoupled Head (Ge et al., 2021). The most important component of the Selective Kernel Attention Decoupled Head is the Selective Kernel Attention. It utilizes multiple

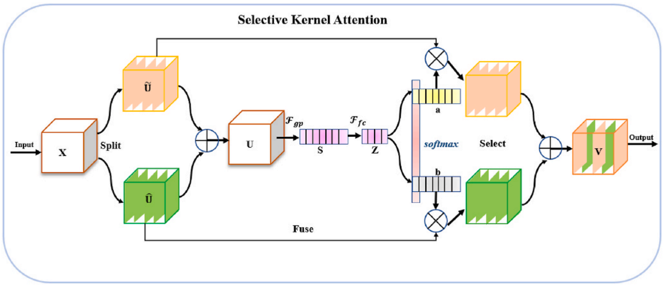

Fig. 8. The structure of Selective Kernel Attention.

parallel convolutional branches, each employing convolution kernels of different sizes, to extract features at various scales. Subsequently, it adaptively adjusts the feature weights of different scales through a self-attention mechanism, ultimately achieving multiscale fusion of the key features of cow mounting behaviour across different scales. This mechanism enables EFF-YOLO to better detect cow mounting behaviourin drastic scale changes environments. The structure of Selective Kernel Attention Decoupled Head is shown in Fig. 6, and the structure of Selective Kernel Attention is shown in Fig. 8.

Selective Kernel Attention implements Selective Kernel convolution through three operations—***Split***, ***Fuse***, and ***Select***, as shown in Fig. 8.

2.5.2.2.1. Split. The split operation is a transformation block consisting of group convolution, batch normalization (Ioffe & Szegedy, 2015, June) and RELU (Glorot et al., 2011) functions.

2.5.2.2.2. Fuse. To achieve adaptive RF adjustment, Selective Kernel Attention uses gates to control information flow. Therefore, the gate needs to fuse all branch information. First, the Selective Kernel Attention fuse the results of multiple branches by summing, as shown in Eq. (1):

$$\boldsymbol{U} = \widetilde{\boldsymbol{U}} + \widehat{\boldsymbol{U}} \tag{1}$$

In the formula, $\widetilde{\boldsymbol{U}}$ is generated by applying a 3 × 3 convolutional kernel to the input feature maps, and $\widehat{\boldsymbol{U}}$ is generated by applying a 5 × 5 convolutional kernel to the input feature maps.

Global information is then obtained by global average pooling, thereby generating channel statistics $\boldsymbol{S}$ as shown in Eq. (2):

$$\boldsymbol{S} = \mathscr{F}_{gp}(\boldsymbol{U}) = \frac{\mathbf{1}}{\boldsymbol{H} * \boldsymbol{W}} \sum_{i=1}^{H} \sum_{j=1}^{W} \boldsymbol{U}(\boldsymbol{i}, \boldsymbol{j}) \tag{2}$$

where $\mathscr{F}_{gp}$ is the global average pooling and H and W refer to the height and width of a feature map, respectively.

Furthermore, a compact feature $\boldsymbol{Z}$ is created by fully connected layers for adaptive and precise selection guidance as shown in Eq. (3):

$$\boldsymbol{Z} = \mathscr{F}_{fc}(\boldsymbol{S}) = \delta(\beta(\boldsymbol{W})) \tag{3}$$

Among them, δ is the RELU function, B is batch normalization and $\mathscr{F}_{fc}$ refers to the fully connected operation.

2.5.2.2.3. Select. Under the guidance of $\boldsymbol{Z}$, Selective Kernel Attention uses cross-channel soft attention to select information adaptively at different spatial scales. Specifically, the softmax operator was applied to the channel numbers as shown in Eq. (4):

$$\boldsymbol{a} = \frac{e^{AZ}}{e^{AZ} + e^{BZ}}, \boldsymbol{b} = \frac{e^{BZ}}{e^{AZ} + e^{BZ}} \tag{4}$$

where $\boldsymbol{a}$, $\boldsymbol{b}$ denote the soft attention vectors of eU and bU, respectively, and A and B denote the transpose matrices corresponding to z. Feature map $\boldsymbol{V}$ was obtained through the attention weights on each kernel, as shown in Eq. (5):

$$\boldsymbol{V} = \boldsymbol{a} * \widetilde{\boldsymbol{U}} + \boldsymbol{b} * \widehat{\boldsymbol{U}}, \boldsymbol{a} + \boldsymbol{b} = \mathbf{1} \tag{5}$$

2.6. Implementation details

This study implemented IATEFF-YOLO in PyTorch1.7.1, and all experimental models were trained and tested on an NVIDIA GTX3090 GPU. During the model training process, using pretrained weights can speed up model training, save training time, ensure model training results, and improve model generalization performance. For the IAT network, the pretraining weights on the LOL dataset (Wei et al., 2018) were adopted, whereas for EFF-YOLO network, the pretraining weights obtained by YOLOX-S training on the COCO dataset (Lin et al., 2014) were adopted.

The input image size was set as 640 × 640. Grid Search was chosen as the algorithm for parameter selection. In this study, the SGD optimizer was chosen for training, and the initial learning rate was set to 0.01. The number of epochs is set as 300. The early stop strategy ensures the best detection accuracy of the model and avoids overfitting. After training for 45 epochs, the early stop strategy was adapted, and training was continued after adjusting the hyperparameters. A larger batch size can reduce the training time and improve the detection accuracy of the model. To balance the computing resources and time, the batch size was set to 32. Cross-entropy loss was used for calculating classification loss, smooth L1 loss is used for localization loss, and binary cross-entropy loss was used for the confidence loss. The Random Shuffle cross-validation technique was used to evaluate model performance.

2.7. Evaluation metrics

To verify the model performance of IATEFF-YOLO, the following four indicators were used to evaluate the proposed model: mean Average Precision (mAP), Precision, Recall and Accuracy. Precision is an effective metric for evaluating the accuracy of a model in predicting positive samples (Zhou X et al., 2020). Recall measures the ability of a model to detect all relevant targets (Zhou C et al., 2023). The mAP combines precision and recall, providing a comprehensive evaluation of model performance (Maxwell A. E. et al., 2021). Accuracy is a commonly used evaluation metric for measuring the accuracy of model predictions. As cows enter oestrus, they frequently engage in mounting behaviour, and each mounting lasts for an average of only 4.0 s. Therefore, the model must be able to quickly detect cow mounting behaviour (Jiang, 2002). Detection speed is also considered an additional evaluation indicator.

2.7.1. Precision

Precision refers to the proportion of cow mounting images detected by the model from all test sets to all cow mounting images in the test set, and can also reflect the model's ability to avoid false detections. The Precision is calculated as shown in Eq. (6):

$$\boldsymbol{Precision} = \frac{TP}{TP + FP} \times 100\% \tag{6}$$

where TP is the number of cow mounting behaviours that are correctly detected in the image and FP is the number of non-cow mounting behaviours detected as mounting behaviours in the image.

2.7.2. Recall

The recall represents the model's ability to correctly detect the cow mounting result based on the actual detection of the cow mounting result, and it focuses more on the missed detection rate. The recall is calculated as shown in Eq. (7):

$$\boldsymbol{Recall} = \frac{TP}{TP + FN} \times 100\% \tag{7}$$

where FN is the number of cow mounting behaviours identified as non-mounting behaviours in the image.

Table 1
Performance of different object detection models on the Test.

Methods	Precision (%)	Recall (%)	mAP (%)	Speed (f/s)	Accuracy (%)
YOLOX	95.6	95.1	95.4	102.0	95.3
SSD	89.8	88.6	91.2	58.0	90.2
Efficentdet	96.3	95.9	93.1	97.0	96.0
RetinaNet	94.0	95.5	95.8	53.0	94.6
DETR	84.5	86.7	85.7	27.8	85.9
Faster RCNN	90.3	89.2	89.7	5.0	89.6
IATEFF-YOLO	**97.8**	**99.4**	**99.3**	**102.0**	**99.1**

2.7.3. mAP

mAP is the mean Average Precision, which can accurately measure the overall performance of IATEFF-YOLO. A higher mAP value also represents a stronger ability of the model to detect cow mounting behaviour, and is calculated as shown in Eq. (8):

$$\boldsymbol{mAP} = \int_0^1 P(R)dR \tag{8}$$

where P is Precision, and R is Recall.

The overall accuracy refers to the proportion of correctly classified examples out of the total number of examples, and it is calculated using Eq. (9):

$$\boldsymbol{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{9}$$

where true negatives (TN) refer to the number of samples that the model correctly identifies as not being instances of mounting cows (i.e., the model predicts that they are not mounting cows, and they actually are not mounting cows).

3. Results

3.1. Comparison with mainstream object detection methods

Mainstream object detection algorithms have already been deeply integrated into various ranch management tasks, with numerous methods specifically focused on livestock behaviour detection (Chen et al., 2021). A direct comparison with these algorithms serves as a proof to the adaptability of this proposed model for ranch management. This study compares methods that are frequently employed in livestock behaviour detection tasks, including Faster R–CNN (Ren et al., 2015), SSD (Liu et al., 2016), RetinaNet (Lin et al., 2017), EfficientDet (Tan et al., 2020), DETR (Carion et al., 2020), and YOLOX (Ge et al., 2021) models. The performance of the mainstream object detection methods on the test set is presented in Table 1.

Compared with mainstream object detection methods, IATEFF-YOLO achieved the best results. Additionally, it has been proven that IATEFF-YOLO can be efficiently applied in pasture management. YOLOX as the baseline model, compared with IATEFF-YOLO, Precision, Recall, mAP and Accuracy lag. Since many cow mounting behaviours occur in complex monitoring scenarios, such as nighttime environments, the model must be able to adapt to these complex scenarios and detect cow mounting behaviours even during the night. However, YOLOX has not considered the challenges faced in practical applications, which is a disadvantage in detecting cow mounting behaviours. Compared with IATEFF-YOLO, RetinaNet and SSD also lag in Precision, Recall, mAP and Accuracy indicators, respectively.

IATEFF-YOLO far exceeded Faster R–CNN. This study found that Faster R–CNN was the slowest among the algorithms, with a speed of only 5.0 f s^{-1}. In ranch management, real-time detection of cow mounting behaviour is essential, so this kind of algorithms is difficult to be suitable for deployment and application within the ranch management system. Although Efficientdet has an efficient feature fusion method (BiFPN), it lags IATEFF-YOLO in all evaluation metrics. The mounting features of cows at far distance are relatively small, requiring greater weights to be assigned to small-scale features in order to achieve efficient fusion of the key mounting features of cows. This necessitates that the model possesses the ability to fuse features adaptively. Under the same training time, the performance of DETR did not meet expectations. Compared with IATEFF-YOLO, DETR still lags in precision, recall, mAP, and accuracy indicators. Because the test set contained many cow mounting images with drastic scale changes and nighttime low-light mounting images, it could not accurately obtain cow mounting features to detect cow mounting. Therefore, it is very important to make targeted improvements to the detection model for nighttime and drastic scale changes in surveillance images caused by different distances between cows and the camera. Compared with mainstream object detection algorithms, the cow mounting behaviour detection algorithm deployed in the pasture management system should balance detection speed and make targeted improvements in response to the problems faced in practical applications.

Table 2
Performance of different object detection models on Test-night.

Methods	Precision (%)	Recall (%)	mAP (%)	Speed (f/s)	Accuracy (%)
SSD	85.7 (↓4.1)	80.6 (↓8.0)	81.8 (↓7.4)	58.0	82.3 (↓7.9)
Efficentdet	85.3 (↓11.0)	60.4 (↓35.5)	74.4 (↓18.7)	97.0	70.9 (↓25.1)
DETR	74.5 (↓10.0)	72.1 (↓14.6)	72.6 (↓14.1)	12.0	72.7 (↓13.2)
RetinaNet	74.6 (↓15.4)	75.4 (↓20.1)	72.3 (↓23.5)	53.0	74.7 (↓19.9)
YOLOX-S	92.3 (↓3.3)	93.1 (↓2.0)	92.8 (↓2.6)	102.0	92.5 (↓2.8)
IATEFF-YOLO	**97.8**	**99.4**	**99.3**	**102.0**	**99.1**

3.2. Comparison of nighttime detection capabilities

Since cows' mounting behaviour frequently occurs at night, a comparison of nighttime detection capabilities can validate IATEFF-YOLO's enhanced efficiency in detecting cows' mounting behaviour under nighttime conditions, providing data support for subsequent algorithm users to analyse this behaviour during the night, and ultimately contributing to the development of smart pastures. To better evaluate the detection ability of the model in the nighttime environment, a test set called Test-night, which was entirely composed of cow mounting tests in the nighttime environment, was selected as the test set. This experiment selects a method that performs better in the test and adds it to the nighttime detection ability test. The performance of each method on the test-night test set is presented in Table 2.

Except for the IATEFF-YOLO model, the detection performance of the other methods dropped significantly. The mAP of EfficentDet, DETR, and RetinaNet decreased by more than 10%. One of the main reasons for the decrease in detection performance is that the design of these methods has not been optimised for detecting cow mounting behaviour at night. The detection accuracy of the IATEFF-YOLO detection network was not affected by the number of night cow mounting images. This study considered the difficulty of night cow mounting detection when designing IATEFF-YOLO. Therefore, at the beginning of the design, the detection model integrated an IAT night enhancement network. The results of the experiment also prove that the IAT can be effectively used for cow mounting detection at night. IATEFF-YOLO presents a novel solution for ranch management to detect cows mounting behaviour during the nighttime, enabling a reduction in the consumption of human resources at nighttime and lowering operational costs of the ranch.

To demonstrate the IAT's ability to enhance low-light images at

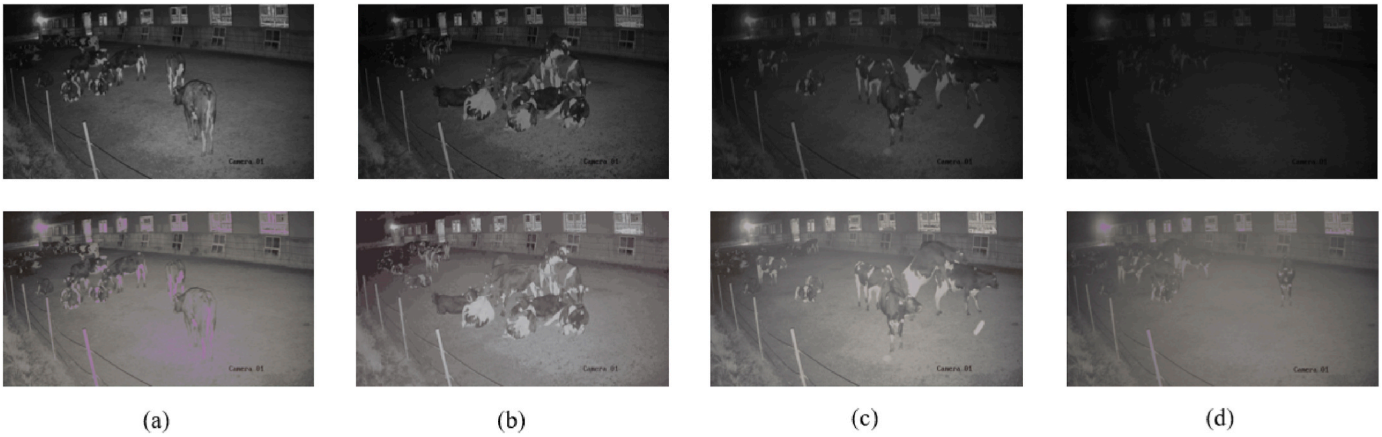

Fig. 9. Examples of test enhancement are shown. (a) and (b) are images of cows mounted under moderate low light. (c) and (d) are the images of cows mounted in severe low light.

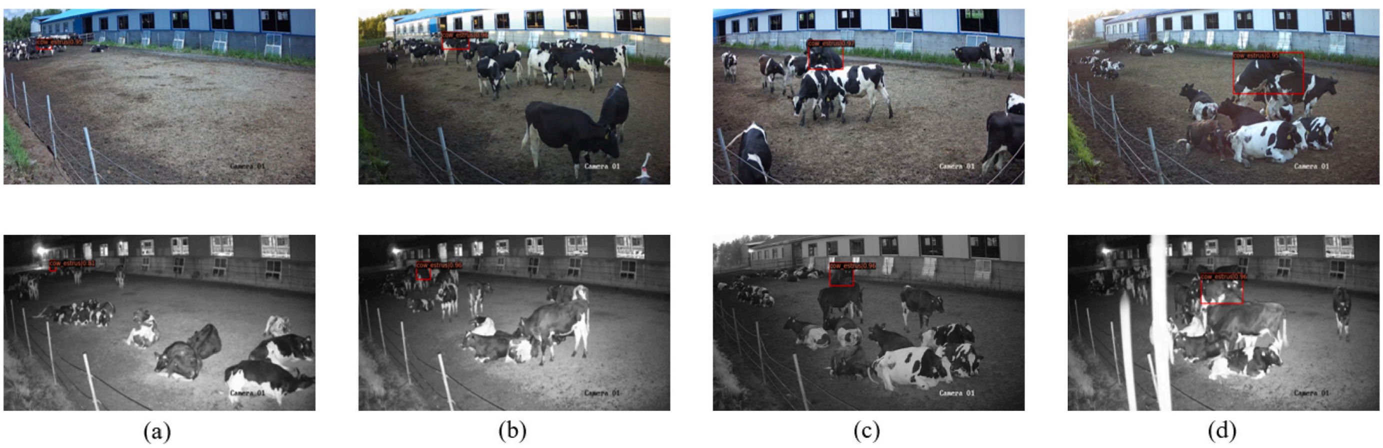

Fig. 10. Examples of scale change detection results. (a) and (b) are the far-distance detection results. (c) are the medium-distance detection results. (d) are the close-distance detection results.

Table 3
Comparison of IATEFF-YOLO and existing cow mounting detection methods.

Methods	Test sets (pieces)	Precision (%)	Recall (%)	mAP (%)	Speed (f/s)	Accuracy (%)
Guo et al. (2019)	949	–	95.8	–	6.9	90.9
Wang, Bai, et al. (2022)	675	94.4	89.8	94.3	61.0	–
Wang et al. (2022)	518	–	–	97.7	50.3	–
NOE et al. (2020)	672	97.0	99.0	–	–	97.0
IATEFF-YOLO	**1710**	**97.8**	**99.4**	**99.3**	**102.0**	**99.1**

night, the enhancement results of images of cows mounting are selected under different low-light levels, as shown in Fig. 9.

Fig. 10 shows the enhancement effect of the model cow mounted images in a night environment at different degrees of darkness. The higher the dimness of the image, as shown in Fig. 9(a) and (b), the better the enhancement effect of the IAT. When the installation image is extremely low-light, as shown in Fig. 9(c) and (d), the IAT can also be enhanced so that the subsequent detection network can obtain the mounting features and ensure the accuracy of model detection.

3.3. Comparison with previous cow mounting detection methods

Comparing with previous cows mounting detection methods to proof model practicality and effectiveness. Furthermore, conducting a thorough analysis of the results enables the offering of technical guidance to assist ranch managers in selecting the most suitable algorithms, thereby enhancing operational efficiency and overall management. A previous detection method for cow mounting with good performance was selected for comparison. Table 3 presents a comparison between IATEFF-YOLO and previous cow mounting detection models.

Among the methods in Table 3, IATEFF-YOLO has the largest test set, which contains a large number of nighttime images and drastic scale change images. NOE et al. (2020) did not explicitly state where the experimental data were collected, whereas the experimental data in the papers by Guo et al. (2019), Wang, Bai, et al. (2022), and Wang et al. (2022b) were collected from outdoor exercise areas. From the images shared by NOE et al. (2020), it appears that the experimental scenarios were single and only involved close-distance scenarios of cow mounting behaviour. The environmental scenarios used by Guo et al. (2019) included only daytime outdoor exercise areas. The environmental scenarios in Wang, Bai, et al. (2022) and Wang et al. (2022b) are consistent with those in this study, all of which collected data from outdoor activity fields under round-the-clock conditions.

The method of Guo et al. (2019) is based on machine learning to detect the mounting behaviour of cows, which detection speed is only 6.9f/s. Due to the complexity of the machine learning method, it is no

Table 4
Impact of each component conducted on the Test dataset.

Methods	mAP (%)	GFLOPs
YOLOX-S	95.4	26.8
YOLOX-S + CSPCSPP	96.2 (↑0.8)	27.5
YOLOX-S + CSPCSPP + SK Decoupled Head (EFF-YOLO)	97.1 (↑1.7)	33.3
IAT + EFF-YOLO (IATEFF-YOLO)	99.3 (↑3.9)	105.8

longer able to meet the real-time requirements of ranch management. Wang, Bai, et al. (2022) used an improved YOLOV5 model that was optimised for multiscale applications and achieved excellent performance. However, owing to the particularity of cow mounting behaviour at night, the model optimisation must consider nighttime detection. IATEFF-YOLO focuses on night detection. The detection speed of the fusion YOLOV5n and channel pruning algorithm proposed by Wang et al. (2022b) was only 50.3 f/s. The detection speed of IATEFF-YOLO proposed in this study was as high as 102.0 f/s, which is twice that of the method proposed by Wang et al. Because the duration of cow mounting is relatively short, when deploying the model in a system to detect mounting cows, the detection speed of the model must be very fast to account for the time consumed by data transmission and to meet the needs of practical applications. In this case, IATEFF-YOLO is more suitable for detecting mounting cows. The method based on machine learning proposed by NOE et al. (2020) reported a recall of 99.0%; however, it relies excessively on manual feature extraction, and the detection process is multistage, which significantly affects the detection efficiency. Compared to previous methods, this study found that most mounting detection methods that rely on machine learning or multistage methods are very slow. Although it exhibits good accuracy, it is unsuitable for deployment in production environments.

The model's exceptional ability to adapt to severe scale changes significantly enhances detection accuracy, bolsters its adaptability and robustness in practical applications, thereby providing better support for ranch management and ultimately improving production efficiency. Because severe scale changes are a common phenomenon in cow mounting detection, this study used a visual method to demonstrate the detection capabilities of IATEFF-YOLO under severe scale changes; some of the test results are shown in Fig. 10.

When the cow mounting behaviour occurs at a relatively far distance from the camera, the network captures only small-scale features of cow mounting, which requires the network to extract key features for detecting cow mounting behaviour from small-scale features. However, when the cow mounting behaviour occurs at a medium or close distance from the camera, the network can obtain the large-scale key features of cow mounting. This requires the detection model to efficiently fuse features from different scales. IATEFF-YOLO possesses a structure that efficiently fuses features of different scales, enabling the accurate fusion of key features of cow mounting behaviour regardless of whether it is faced with large-scale or small-scale cow mounting behaviour features, providing feature support for the subsequent detection of mounting cows, as shown in Fig. 10.

3.4. Analysis of the contribution of different components

Analysing the contributions of different modules can enable ranch managers to clearly understand the design principles of the algorithm, which is of great significance for error analysis and result interpretation in practical applications. The impact of each component on the model's performance is presented in Table 4. GFLOPs are evaluation metrics used to measure the computational complexity of deep learning models and are commonly used in module contribution evaluation. Therefore, the GFLOPs evaluation metric has been added to Table 4.

The YOLOX-S feature fusion method changed SPP to CSPCSPP, mAP increased from 95.4% to 96.2%, and GFLOPs increased from 26.8 to 27.5 and enabled the network to better integrate the features of cow mounting behaviour in environments with severe scale changes. This feature fusion method allows the backbone of IATEFF-YOLO to obtain more features than the backbone of YOLOX-S, which is beneficial for the other components.

After the addition of the SK decoupled head, the mAP increased from 96.2% to 97.1%. This can make the detection head pay more attention to the mounting behaviour of cows, enhance the detection capability of the detection head in environments with severe scale changes, and improve the upper limit of the ability of IATEFF-YOLO to detect the mounting behaviour of cows.

The integration of IAT and the EFF-YOLO mAP increased from 97.1% to 99.3%, whereas the GFLOPs increased from 33.3 to 105.8. IAT contributes the most to the performance improvement of the model. This provides reference for the designers of ranch system algorithms. If the practical application scenarios involve nighttime low-light condition, it is necessary to integrate low-light enhancement framework like IAT.

4. Discussion

4.1. Analysis of the impact of complex illumination on model detection

Since cow mounting behaviour often occurs in outdoor sports areas where illumination conditions are complex. These conditions generally encompass the intricacies of varying light intensities during both day and night, as well as the uneven distribution of light manifested in forms

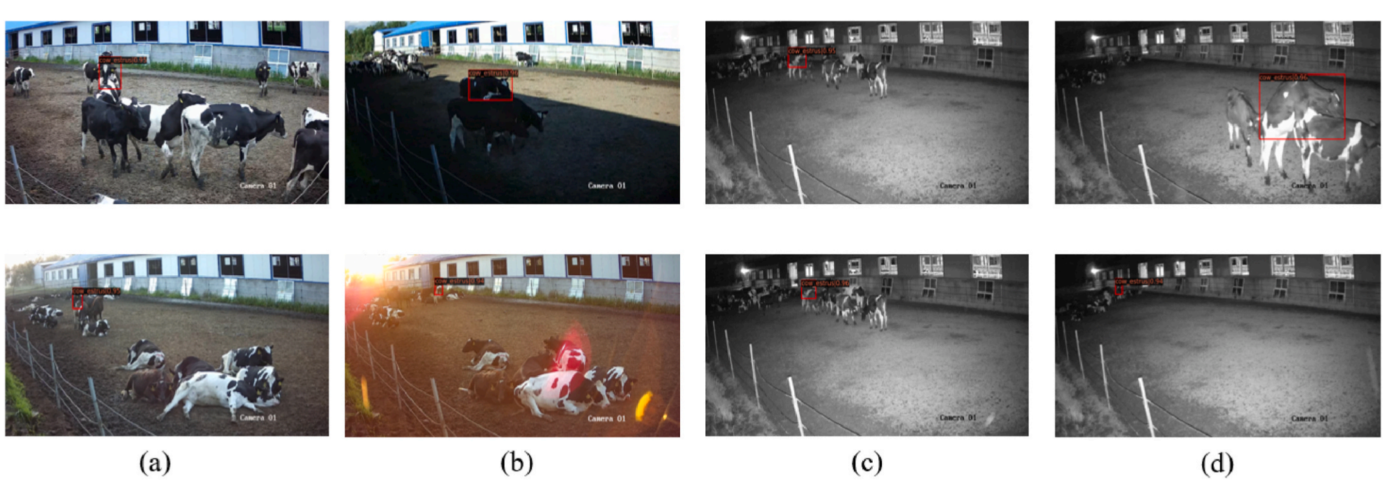

Fig. 11. Examples of complex illumination results. (a) is the detection result under normal light. (b) is the detection of results under complex illumination. (c) and (d) are the images of detected results under nighttime low light.

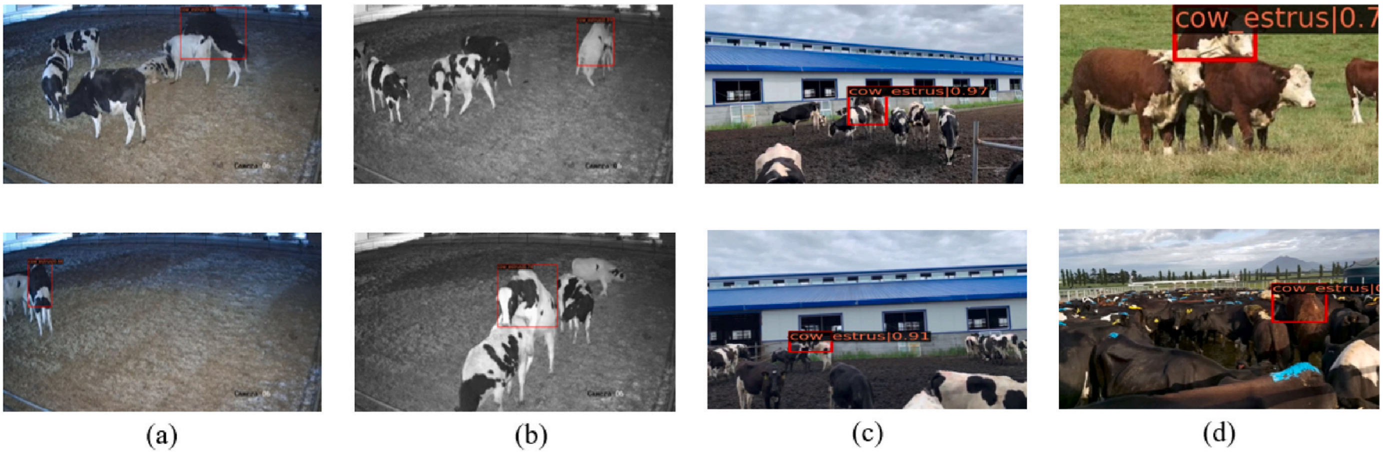

Fig. 12. Examples of different collection methods and cows breeds detection results. (a) and (b) are the detection results after adjusting the tilt angle of the surveillance camera. (c) are the detection results of mobile phones captured. (d) are the detection results of Ayrshire cows.

such as shadows, veiling glare, and ghosting effects. Such complexity can result in the loss or blurring of details in cow mounting images, which can significantly affect the detection of cow mounting behaviour. Therefore, analysing the detection performance of the model under complex lighting conditions is of great significance. Fig. 11 shows the detection results of IATEFF-YOLO for complex illumination scenarios.

As shown in Fig. 11(a), IATEFF-YOLO performed well in a normal light environment and, surprisingly, could still accurately detect the mounting behaviour of cows in an occluded environment. Fig. 11(b) shows the detection results under complex illumination conditions. The shadows caused by the angle of light illumination that covers the features of the cows or the veiling glare and ghosting caused by direct light hitting the surveillance camera result in poor image quality of cow mounting images. IATEFF-YOLO exhibited powerful detection capabilities.

Fig. 11(c) and (d) show the detection results in the nighttime environment. IATEFF-YOLO also achieved accurate detection of mounting cows in a nighttime environment, providing a new solution for detecting cow mounting behaviour in complex illumination environments.

4.2. Analysis of the model robustness of different collection methods and cows breeds

Analysing the robustness of a model holds significant value for its practical application, as it reflects the model's reliability, stability, and generalisation ability in application environments. Therefore, to further evaluate the robustness of IATEFF-YOLO, a test-robust test set is employed for visualisation testing.

Despite the change in the tilt angle of the surveillance camera, IATEFF-YOLO still accurately detected mounting cows, as shown in Fig. 12(a) and (b). When IATEFF-YOLO was used to detect mounting cows photographed by a mobile phone, it still exhibited good detection performance. When faced with non-Holstein mounting cows, it still showed good generalisation ability. This is inherently linked to the ability of IATEFF-YOLO to accurately extract key features of cow mounting behaviours. This demonstrates the value of IATEFF-YOLO for practical application.

4.3. Analysis of cows mounting behaviour statistical application

This study also randomly selected a day's surveillance video and provided it to IATEFF-YOLO to simulate the application scene on commercial farms; the number of cows mounted was determined using IATEFF-YOLO. On the experimental farm, the number of Holstein cows in the exercise area was not fixed. When the barns were cleaned, all 107 heifers remained in the outdoor exercise area. However, during feeding, all heifers were driven back to the barns. All the statistical analyses in this section only consider the changes in the regularity of cow mounting behaviour across different time periods, and the influences of factors such as air temperature and solar radiation have not been considered in this study. The statistical results are shown in Fig. 13.

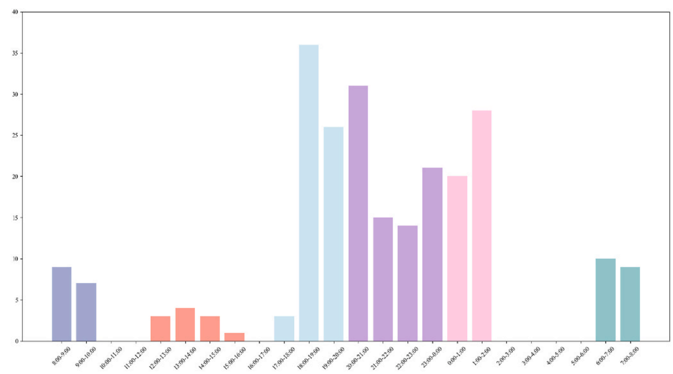

Fig. 13. Statistical results of cow mounting times during one day.

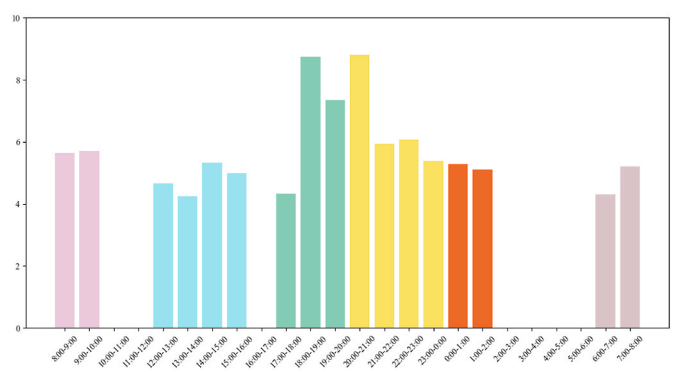

Fig. 14. Average duration of the cows' mounting behaviour in each time period.

This study found that the number of cow mounting behaviours increased significantly starting in the evening and continued until the first half of the night. The number of mounting times was mainly concentrated between 18:00 and 2:00 on the next day, accounting for approximately 79.6%. This confirms that cow mounting behaviour

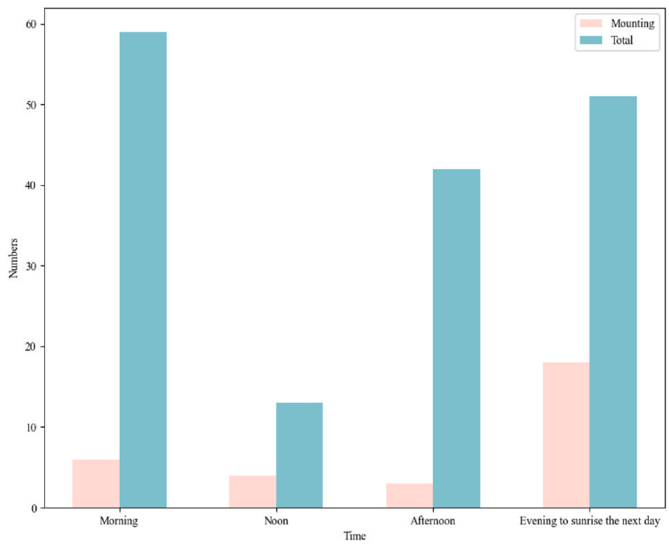

Fig. 15. Number of mounting cows and the total number of cows in the outdoor exercise area at different time.

occurs more often at night than during the day.

This study also calculated the average duration of cow mounting behaviour within each time interval, as shown in Fig. 14.

As shown in Fig. 14, as the frequency of cow mounting behaviour increased during certain time intervals, the average duration of cow mounting behaviour also increased correspondingly. This provides inspiration for designers of cow mounting behaviour detection systems. To achieve real-time detection of the cow mounting behaviour, the model detection time and communication time between the system modules should not exceed 5 s.

This paper also conducted statistical analyses of the percentage of mounting cows at different times. This study conducted three days of observations and found that in the morning (6:00–12:00), there were approximately 59 Holstein heifers; around noon (12:00–14:00) there were approximately 13 heifers; in the afternoon (14:00–18:00) there were approximately 42 Holstein heifers; and from evening to sunrise the next day (18:00–6:00), there were approximately 51 Holstein heifers. The number of mounted cows and total number of cows corresponding to that period are shown in Fig. 15.

The percentages of mounting cows at different time as shown in Fig. 16.

The reason for the relatively high percentages during noon is not due to a higher frequency of mounting behaviours, but rather to the fact that there are fewer cows in the exercise area. Combining Figs. 13, 15 and 16 provides new insights into cow mounting management for pastures. Branch-breeding managers can focus their efforts from 6:00 a.m. to 10:00 a.m. and from 6:00 p.m. to 2:00 a.m. the next day. This can also enhance the efficiency of oestrus management for cows on the ranch, allowing for more effective monitoring of oestrus with limited manpower. Consequently, the empty period of the cows was shortened, thereby improving the economic efficiency of the ranch.

4.4. Analysis of IATEFF-YOLO application limitations

Although IATEFF-YOLO can accurately and quickly detect cow mounting behaviour, it has limitations. Fig. 17 shows some examples of detection failures. The upper half of Fig. 17 shows the ground truth, and the lower half shows the detection results of the model.

Since it is difficult to extract the mounting features of dairy cows in distant, seriously occluded environments, the occlusion between dairy cows exacerbates the lack of cow mounting features, as shown in Fig. 17 (a) and (b). The backbone of IATEFF-YOLO is unable to capture global information, leading to failure in detecting the mounting behaviour of dairy cows in distantly occluded environments.

Due to the decrease in temperature at night, water vapour condensation can occur, and when the surface temperature begins to rise, fogging can easily occur, as shown in Fig. 17(c) and (d). Aggravated lack of cow mounting features under foggy conditions for detecting mounting cows. Moreover, this study also found that failed cases of detecting the mounting behaviour of dairy cows in distant seriously occluded environments or distant foggy environments often involved dairy cows with back patterns similar to the environment. This is also one of the reasons why IATEFF-YOLO was unable to detect them.

Currently, IATEFF-YOLO does not have an identity module for identifying cows; therefore, it cannot distinguish whether a cow's

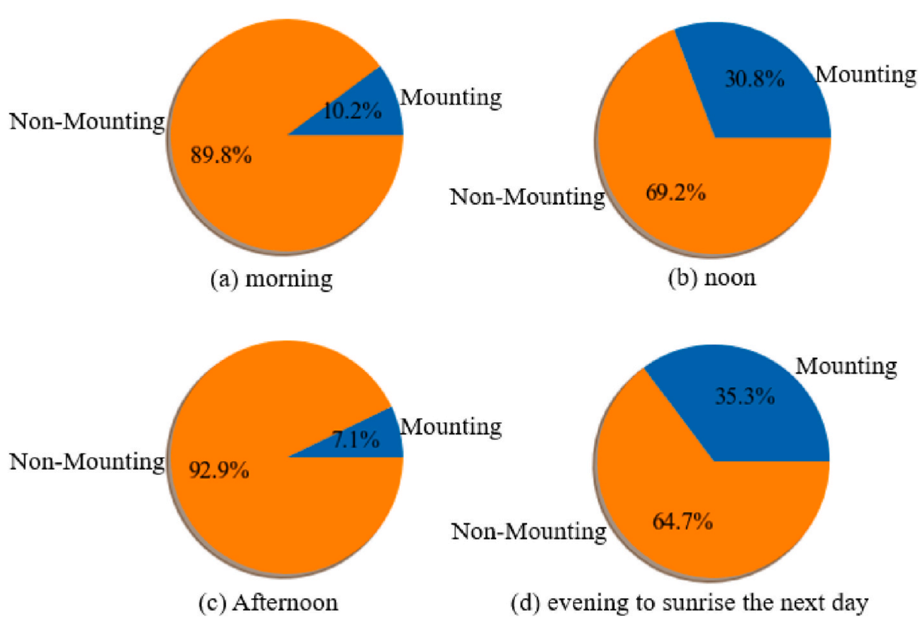

Fig. 16. Percentage of dairy cows mounted in the outdoor exercise area at different times compared to the total number of dairy cows. (a) is the percentage of cows that mount in the morning (6:00–12:00) to the total number of cows. (b) is the percentage of cows that mount at noon (12:00–14:00) to the total number of cows. (d) is the percentage of cows that mount at noon (12:00–14:00) to the total number of cows.

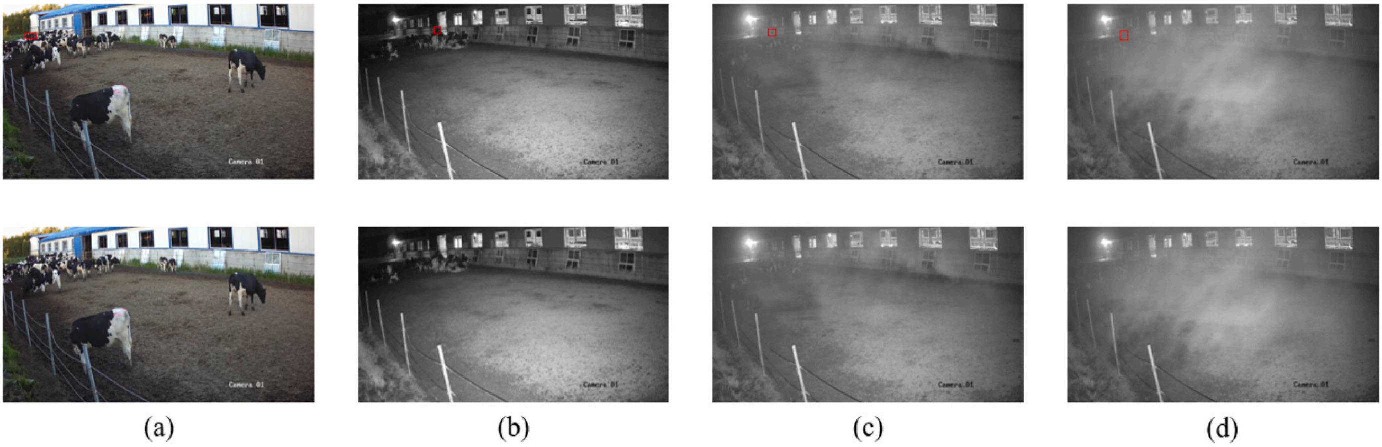

Fig. 17. Examples of detection failures. (a) and (b) is far-distance occlusion detection results. (c) and (d) are far-distance fog detection results.

mounting behaviour originates from the same cow. Identifying the identity of cows relies on recognition by breeding personnel on farms. In future work, the cow identity module (Zheng et al., 2016) will be integrated with a cow mounting behaviour detection model. The impact of factors such as air temperature and solar radiation on the regularity of cow mounting behaviour will also be given emphasis in the future work of this study.

5. Conclusion

In this work, a novel cow mounting detector is proposed named IATEFF-YOLO. IATEFF-YOLO consists of Illumination Adaptive Transformer and EFF-YOLO detector. The experimental results show that IATEFF-YOLO can effectively detect the mounting behaviour of cows in the case of nighttime and drastic scale changes in surveillance images caused by different distances between cows and camera, with high accuracy and detection speed in detecting cow mounting. On test set, IATEFF-YOLO has a mean Average Precision of 99.3% and a detection speed of 102.0 f s^{-1}, which still achieves better detection results compared to other methods, so this proves that the IATEFF-YOLO model detection performance. The non-contact detection model of cow mounting proposed in this paper will not cause stress response to cows and is more beneficial to animal welfare.

CRediT authorship contribution statement

De Li: Writing – original draft, Visualization, Validation, Software, Methodology, Data curation, Conceptualization. **Baisheng Dai:** Writing – review & editing, Visualization, Resources, Funding acquisition, Formal analysis, Data curation. **Yanxing Li:** Writing – review & editing. **Peng Song:** Validation, Resources. **Xin Dai:** Resources. **Yongqiang He:** Resources. **Huixin Liu:** Resources, Data curation. **Yang Li:** Funding acquisition, Data curation. **Weizheng Shen:** Writing – review & editing, Funding acquisition, Formal analysis, Data curation.

Declaration of competing interest

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.

Acknowledgments

This work was supported in part by the National Key Research and Development Program of China (Grant No. 2023YFD2000700), in part by the Key Research and Development Program of Heilongjiang Province (Grant No. 2022ZX01A24), in part by the National Natural Science Foundation of China (Grant No. 32072788), in part by the Modern Agricultural Industry Technology Collaborative Innovation Promotion System Construction Project of Heilongjiang Province (the Letter of Department of Agriculture and Rural Affairs of Heilongjiang Province (2021) No. 1492), and in part by the earmarked fund for CARS (Grant No. CARS-36).

References

Bochkovskiy, A., Wang, C. Y., & Liao, H. Y. M. (2020). Yolov4: Optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934 https://doi.org/10.48550/arXiv.2004.10934.

Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., & Zagoruyko, S. (2020). End-to-end object detection with transformers. In *European conference on computer vision* (pp. 213–229). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-58452-8_13.

Chae, J. W., & Cho, H. C. (2021). Identifying the mating posture of cattle using deep learning-based object detection with networks of various settings. *Journal of Electrical Engineering & Technology, 16*(3), 1685–1692. https://doi.org/10.1007/s42835-021-00701-z

Chen, C., Zhu, W., & Norton, T. (2021). Behaviour recognition of pigs and cattle: Journey from computer vision to deep learning. *Computers and Electronics in Agriculture, 187*, Article 106255. https://doi.org/10.1016/j.compag.2021.106255

Chung, Y., Choi, D., Choi, H., Park, D., Chang, H. H., & Kim, S. (2015). Automated detection of cattle mounting using side-view camera. *KSII Transactions on Internet and Information Systems (TIIS), 9*(8), 3151–3168. https://koreascience.kr/article/JAKO201536553536306.pdf.

Cui, Z., Li, K., Gu, L., Su, S., Gao, P., Jiang, Z., … Harada, T. (2022). You only need 90k parameters to adapt light: A light weight transformer for image enhancement and exposure correction. *arXiv preprint arXiv:2205.14871*. https://doi.org/10.48550/arXiv.2205.14871

Esslemont, R. J., Glencross, R. G., Bryant, M. J., & Pope, G. S. (1980). A quantitative study of pre-ovulatory behaviour in cattle (British Friesian heifers). *Applied Animal Ethology, 6*(1), 1–17. https://doi.org/10.1016/0304-3762(80)90090-5

Ge, Z., Liu, S., Wang, F., Li, Z., & Sun, J. (2021). Yolox: Exceeding yolo series in 2021. arXiv preprint arXiv:2107.08430 https://doi.org/10.48550/arXiv.2107.08430.

Glorot, X., Bordes, A., & Bengio, Y. (2011). Deep sparse rectifier neural networks. In *Proceedings of the fourteenth international conference on artificial intelligence and statistics* (pp. 315–323). JMLR Workshop and Conference Proceedings. http://proceedings.mlr.press/v15/glorot-11a.

Guo, Y., Zhang, Z., He, D., Niu, J., & Tan, Y. (2019). Detection of cow mounting behavior using region geometry and optical flow characteristics. *Computers and Electronics in Agriculture, 163*, Article 104828. https://doi.org/10.1016/j.compag.2019.05.037

He, K., Zhang, X., Ren, S., & Sun, J. (2015). Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 37*(9), 1904–1916. https://doi.org/10.1109/TPAMI.2015.2389824

Higaki, S., Okada, H., Suzuki, C., Sakurai, R., Suda, T., & Yoshioka, K. (2021). Estrus detection in tie-stall housed cows through supervised machine learning using a multimodal tail-attached device. *Computers and Electronics in Agriculture, 191*, Article 106513. https://doi.org/10.1016/j.compa-g.2021.106513

Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International conference on machine learning* (pp. 448–456). pmlr. http://proceedings.mlr.press/v37/ioffe15.html.

Jiang, H. (2002). Comparison of three cow estrus identification methods. *Chinese Journal of Animal Husbandry, (5)*, 37–38.

Jocher, G. (2021). YOLOv5 release v6.0[EB/OL]. https://github.com/u-ltralytics-/yolov5/rel-eases/tag/v6.0.

Li, X., Wang, W., Hu, X., & Yang, J. (2019). Selective kernel networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 510–519). https://open-access.thecvf.com/content_CVPR_2019/papers/Li_Selective_Kernel_Networks_CVPR_2019_paper.pdf.

Lin, T. Y., Goyal, P., Girshick, R., He, K., & Dollár, P. (2017). Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision* (pp. 2980–2988). https://openaccess.thecvf.com/content_ICCV_2017/papers/Lin_Focal_Loss_for_ICCV_2017_paper.pdf.

Lin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., … Zitnick, C. L. (2014). Microsoft coco: Common objects in context. In *CompSter vision–ECCV 2014: 13th European conference, Zurich, Switzerland, september 6-12, 2014, proceedings, Part V* (pp. 740–755). Springer International Publishing. https://doi.org/10.1007/978-3-319-10602-1_48, 13.

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. (2016). Ssd: Single shot multibox detector. In *Computer vision–ECCV 2016: 14th European conference, Amsterdam, The Netherlands, october 11–14, 2016, proceedings, Part I 14* (pp. 21–37). Springer International Publishing. https://doi.org/10.1007/978-3-319-46448-0_2.

Liu, S., Qi, L., Qin, H., Shi, J., & Jia, J. (2018). Path aggregation network for instance segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 8759–8768). https://openaccess.thecvf.com/content_cvpr_2018/papers/Liu_Path_Aggregation_Network_CVPR_2018_paper.pdf.

Lodkaew, T., Pasupa, K., & Loo, C. K. (2023). CowXNet: An automated cow estrus detection system. *Expert Systems with Applications, 211*, Article 118550. https://doi.org/10.1016/j.eswa.202-2.118550

MacKay, J. R., Deag, J. M., & Haskell, M. J. (2012). Establishing the extent of behavioural reactions in dairy cattle to a leg mounted activity monitor. *Applied Animal Behaviour Science, 139*(1–2), 35–41. https://doi.org/10.1016/j.applanim.2012.03.008

Maxwell, A. E., Warner, T. A., & Guillén, L. A. (2021). Accuracy assessment in convolutional neural network-based deep learning remote sensing studies—Part 1: Literature review. *Remote Sensing, 13*(13), 2450. https://doi.org/10.3390/rs13132450

Mwaanga, E. S., & Janowski, T. (2000). Anoestrus in dairy cows: Causes, prevalence and clinical forms. *Reproduction in Domestic Animals, 35*(5), 193–200. https://doi.org/10.1046/j.14-39-0531.2000.00211.x

Noe, S. M., Zin, T. T., Tin, P., & Hama, H. (2020). Detection of estrus in cattle by using image technology and machine learning methods. In *2020 IEEE 9th global conference on consumer electronics (GCCE)* (pp. 320–321). IEEE. https://doi.org/10.1109/GCCE50665.20-20.9291987.

Pasupa, K., & Lodkaew, T. (2019). A new approach to automatic heat detection of cattle in video. In *Neural information processing: 26th international conference, ICONIP 2019, sydney, NSW, Australia, december 12–15, 2019, proceedings, Part V 26* (pp. 330–337). Springer International Publishing. https://doi.org/10.1007/978-3-030-36802-9_35.

Perez Marquez, H. J., Ambrose, D. J., & Bench, C. J. (2023). Behavioral changes to detect estrus using ear-sensor accelerometer compared to in-line milk progesterone in a commercial dairy herd. *Frontiers in Animal Science, 4*, Article 1149085. https://doi.org/10.3389/fanim.2023.11-49085

Redmon, J., & Farhadi, A. (2018). Yolov3: An incremental improvement. arXiv preprint arXiv:18-04.02767 https://doi.org/10.48550/arXiv.1804.02767.

Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in Neural Information Processing Systems, 28*. https://proceedings.neurips.cc/paper_files/paper/2015/file/14bfa6bb14875e45bba028a21ed38046-Paper.pdf.

Roelofs, J., Lopez-Gatius, F., Hunter, R. H. F., Van Eerdenburg, F. J. C. M., & Hanzen, C. H. (2010). When is a cow in estrus? Clinical and practical aspects. *Theriogenology, 74*(3), 327–344. https://doi.org/10.1016/j.theriogenology.2010.02.016

Roelofs, J. B., Van Eerdenburg, F. J. C. M., Soede, N. M., & Kemp, B. (2005). Various behavioral signs of estrous and their relationship with time of ovulation in dairy cattle. *Theriogenology, 63*(5), 1366–1377. https://doi.org/10.1016/j.theriogenology.2004.07.009

Sumi, K., Zin, T. T., Kobayashi, I., & Horii, Y. (2018). Framework of cow calving monitoring system using a single depth camera. In *2018 international conference on image and vision computing New Zealand (IVCNZ)* (pp. 1–7). IEEE. https://doi.org/10.1109/IVCN-Z.2018.8634738.

Tan, M., Pang, R., & Le, Q. V. (2020). Efficientdet: Scalable and efficient object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 10781–10790). https://openaccess.thecvf.com/content_CVPR_2020/papers/Tan_EfficientDet_Scalable_and_Efficient_Object_Detection_CVPR_2020_paper.pdf.

Tsai, D. M., & Huang, C. Y. (2014). A motion and image analysis method for automatic detection of estrus and mating behavior in cattle. *Computers and Electronics in Agriculture, 104*, 25–31. https://doi.org/10.1016/j.compag.2014.03.003

Van Vliet, J. H., & Van Eerdenburg, F. J. C. M. (1996). Sexual activities and oestrus detection in lactating Holstein cows. *Applied Animal Behaviour Science, 50*(1), 57–69, 1-0.1016/0168-1591(96)01068-4.

Wang, R., Bai, Q., Gao, R., Li, Q., Zhao, C., Li, S., & Zhang, H. (2022a). Oestrus detection in dairy cows by using atrous spatial pyramid and attention mechanism. *Biosystems Engineering, 223*, 259–276. https://doi.org/10.1016/j.biosystemseng.2022.08.018

Wang, J., Bell, M., Liu, X., & Liu, G. (2020a). Machine-learning techniques can enhance dairy cow estrus detection using location and acceleration data. *Animals, 10*(7), 1160. https://do-i.org/10.3390/ani10071160.

Wang, C. Y., Liao, H. Y. M., Wu, Y. H., Chen, P. Y., Hsieh, J. W., & Yeh, I. H. (2020b). CSPNet: A new backbone that can enhance learning capability of CNN. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops* (pp. 390–391). https://open-access.thecvf.com/content_CVPRW_2020/papers/w28/Wang_CSPNet_A_New_Backbone_That_Can_Enhance_Learning_Capability_of_CVPRW_2020_paper.pdf.

Wang, Z., Wang, S., Wang, C., Zhang, Y., Zong, Z., Wang, H., … Du, Y. (2023). A non-contact cow estrus monitoring method based on the thermal infrared images of cows. *Agriculture, 13*(2), 385. https://doi.org/10.3390/agriculture13020385

Wang, Z., Xu, X., Hua, Z., Shang, Y., Duan, Y., & Song, H. (2022b). Lightweight recognition for the oestrus behavior of dairy cows combining YOLO v5n and channel pruning. *Transactions of the Chinese Society of Agricultural Engineering, 38*, 130–140.

Wei, C., Wang, W., Yang, W., & Liu, J. (2018). Deep retinex decomposition for low-light enhancement. *arXiv preprint arXiv:1808.04560*. https://doi.org/10.48550/arXiv.1808.04560

Yusheng, H., Qiang, L., Weilan, W., & Yulan, L. (2020). Identification of cow mounting behavior based on Wi-Fi wireless sensing technology. *Transactions of the Chinese Society of Agricultural Engineering, 36*(18).

Zheng, L., Yang, Y., & Hauptmann, A. G. (2016). Person re-identification: Past, present and future. *arXiv preprint arXiv:1610.02984*. https://doi.org/10.48550/arXiv.1610.02984

Zhou, C., Lu, Z., Lv, Z., Meng, M., Tan, Y., Xia, K., … Zuo, H. (2023). Metal surface defect detection based on improved YOLOv5. *Scientific Reports, 13*(1), Article 20803, 10.103-8/s41598-023-47716-2.

Zhou, X., Lu, P., Zheng, Z., Tolliver, D., & Keramati, A. (2020). Accident prediction accuracy assessment for highway-rail grade crossings using random forest algorithm compared with decision tree. *Reliability Engineering & System Safety, 200*, Article 106931. https://doi.org/10.1016/j.ress.202-0.106931

Zoph, B., Cubuk, E. D., Ghiasi, G., Lin, T. Y., Shlens, J., & Le, Q. V. (2020). Learning data augmentation strategies for object detection. In *Computer vision–ECCV 2020: 16th European conference, glasgow, UK, August 23–28, 2020, proceedings, Part XXVII* (pp. 566–583). Springer International Publishing. https://doi.org/10.1007/978-3-030-58583-9_34, 16.