

Non-contact respiration rate measurement of multiple cows in a free-stall barn using computer vision methods

Hang Shu ^{a,b,c,*}, Jérôme Bindelle ^c, Xianhong Gu ^{a,*}

^a State Key Laboratory of Animal Nutrition, Institute of Animal Sciences, Chinese Academy of Agricultural Sciences, Beijing 100193, China

^b Agricultural Information Institute, Chinese Academy of Agricultural Sciences, Beijing 100086, China

^c AgroBioChem/TERRA, Precision Livestock and Nutrition Unit, Gembloux Agro-Bio Tech, University of Liège, 5030 Gembloux, Belgium



ARTICLE INFO

Keywords:

Precision livestock farming
Animal welfare
Deep learning
Multi-object measurement
Heat stress

ABSTRACT

In cattle, respiration rate (RR) provides researchers and practitioners with valuable physiological information. However, traditional visual observation requires massive labour and is impractical for large-scale commercial farms. Recently developed vision-based RR measurement methods are highly limited to measuring a small number of animals in a controlled environment. Therefore, this paper aimed to propose a vision-based multi-object RR measurement method for dairy cows lying in free stalls. An RGB camera was aimed at the lying zone and was able to cover about 16 stalls. The proposed framework first utilised two YOLOv5-based networks to segment cow instances and detect cow flank objects. Next, an object tracker was used to link the predictions of each cow throughout the video clip. The Lucas-Kanade optical flow was then calculated specifically on the overlapped area of the cow mask and the flank bounding box. Finally, RR was extracted using Fast Fourier Transform. The results show that the proposed method had a precise RR measurement with a correlation coefficient of 0.944, a root mean square error of 5.35 breaths per minute, and an intraclass correlation coefficient of 0.974 when compared to visual observation. The piecewise regression models identified a change in RR when the ambient temperature reached 23.6 °C or the temperature-humidity index reached 72. The corresponding RR thresholds were 60.9 and 60.2 breaths per minute, respectively. Collectively, these results can be used to inform an automated local cooling system, e.g., fans in the lying area. However, more experiments and calibration with data collected using more cost-effective video recording systems are required before this technology can be applied on farms.

1. Introduction

Respiration rate (RR) is a basic vital sign of animals that indicates their physiological state and overall health condition. For example, an abnormal RR in cattle can be used as an alert for the onset of common events of interest, such as heat stress (Shu et al., 2022). However, traditional RR measurement by manually counting flank movements over a set period of time and converting them to breaths per minute (bpm) can be tedious and time-consuming in frontline practice and is also prone to biased results due to the lack of professionals.

In order to solve the abovementioned problems, many precision livestock farming techniques have recently been put forward to achieve an automated RR measurement. These techniques rely on different methodologies to collect respiration-related signals, which are then

processed to determine RR using proper signal processing methods or more advanced deep learning methods. Depending on whether physical contact with animals is required, these methods can be divided into two categories: contact and non-contact methods.

On the one hand, contact methods typically involve placing devices on the cows' bodies to collect respiration-related signals. For example, an MP3 recorder on a halter can collect the acoustic signal of breathing (de Carvalho et al., 2020), while pressure sensors (Atkins et al., 2018) can measure RR relying on chest and abdomen movements. Pressure sensors (Strutzke et al., 2019) or thermistors (Milan et al., 2016) can also be mounted to halters making use of the pressure and temperature difference between inhaled and exhaled air, respectively. More recently, accelerometers (Bar et al., 2019) have been validated to be able to detect small movements during heavy breathing.

* Corresponding authors at: State Key Laboratory of Animal Nutrition, Institute of Animal Sciences, Chinese Academy of Agricultural Sciences, Beijing 100193, China (H. Shu and X. Gu).

E-mail addresses: shuhang@caas.cn (H. Shu), guxianhong@vip.sina.com (X. Gu).

On the other hand, non-contact methods for measuring RR are drawing increased interest since they better meet animal welfare requirements by eliminating the stress and discomfort of cows. In cattle studies, various techniques and devices have been applied to measure RR, such as radar (Tuan et al., 2022), laser (Pastell et al., 2006), infrared cameras (Lowe et al., 2019), and RGB cameras (Fuentes et al., 2021). Radar and laser methods rely on respiration-related movements, while the infrared method relies on the changes in skin or air temperature around nose areas caused by breathing, all of which require expensive devices and thus may be unaffordable for most farms.

Also, respiration-related body movements can be detected more cost-effectively through an RGB camera and processed further by cattle segmentation, optical flow calculation, and statistical analysis to generate RR estimation (Wu et al., 2023; Wu et al., 2020). However, the reported method was only designed for a controlled scenario in which a single or two targeted cows were held in a fixed sideway pose in the field of view. Moreover, respiration-related body movements also result in periodic fluctuations in the brightness of colour channels. Consequently, RR can be directly estimated from the variations in pixel intensity of the RGB channels of certain regions of interest (e.g., chest and abdomen). This method has shown its effectiveness in a multi-animal environment for measuring up to five pigs (Wang et al., 2023) and four cows (Mantovani et al., 2023), but has yet to be tested in cows raised in a larger-scale free-stall pen. In addition, the colour or brightness changes in the skin due to changes in blood volume during respiration cycles can be measured in proper colour channels using the remote photoplethysmography (rPPG) principle. This method, however, requires a closer distance between the camera and the animal in order to capture the subtle respiration-related blood flow changes in veins. rPPG-based method has been used to measure RR for hairless animals like pigs (Jorquera-Chavez et al., 2020). However, in the case of dairy cows, which typically have longer hair, the requirement of a direct view of body areas with less hair (e.g., facial area) by rPPG renders the large-scale measurement in free stalls impossible.

In summary, the existing RR measurement methods for dairy cows require improvement as they are either too expensive, require physical contact with the animal, or are limited to measuring very few animals in a controlled scenario. For a larger-scale application in a practical farm, it is necessary to further integrate certain automated technologies such as cow and regions of interest recognition, multi-object tracking, and RR calculation. Thus, this paper aimed to propose a vision-based multi-object RR measurement pipeline for dairy cows housed in an uncontrolled scenario, i.e., a commercial free-stall barn, and further investigate its feasibility in indicating heat stress. By leveraging deep learning and optical flow techniques, the proposed method was expected to provide a more cost-effective and non-invasive solution for farmers and veterinarians to detect significant events, such as heat stress, in an early stage.

2. Materials and methods

All protocols involving animals were approved by the Experimental Animal Care and Use Committee of the Institute of Animal Sciences, Chinese Academy of Agricultural Sciences (approval number 2016IAS018).

2.1. Experimental design

The field experiment was carried out on a free-stall dairy farm in Beijing, China for 35 days from late August to early October 2017. The location is characterised by a temperate continental monsoon climate with hot and humid summers.

The experimental barn (oriented along the north–south axis) had four pens with a typical 4-row head-to-head design and was covered by a roof. All four pens (15 m × 90 m) were used in this experiment, housing 79, 70, 80, and 60 lactating Holstein-Friesian dairy cows, respectively.

For each pen, ambient temperature (Ta, °C) and relative humidity (RH, %) were measured at an interval of 5 min using three Kestrel 5000 environment meters (accuracy: ± 0.4 °C Ta, ± 1 % RH; Nielsen-Kellerman, Boothwyn, PA, USA). The sensors were fixed at a height of 2 m and were evenly placed at one-fourth, one-half, and three-fourths of the total length of the pen, respectively. The temperature-humidity index (THI) was calculated as per Eq. (1) (NRC, 1971).

$$THI = (1.8 \times Ta + 32) - (0.55 - 0.005 \times RH) \times (1.8 \times Ta - 26) \quad (1)$$

A Sony HDR-CX405 camera (Sony, Tokyo, Japan) was fixed on a telescopic camera stick with an angle of approximately 45° downward. The frame rate of the recordings was set to 30 frames per second (FPS). The camera was placed at a height of 4 m along the feeding alley. The camera was aimed at the lying zone and was able to cover approximately 16 stalls. The reason for focusing on lying cows is that the RR of standing cows is rather difficult to detect due to their extremely subtle breathing movements and frequent accompanying interfering movements. Lying cows, on the contrary, reflect heat stress response to the ambient environment earlier (Pinto et al., 2020), and thus provide a good representative of the herd. The video clips of 10–25 s were captured seven times daily, from 07:30 to 17:30 h, for the four pens sequentially.

2.2. Data preparation

A video frame was randomly extracted from each video clip by using a self-written Python program. This was done to avoid duplicating similar frames since cows had relatively small and slow displacements in the video clips. The extracted frames were in JPG format with a resolution of 1,920 by 1,080 pixels. Low-quality frames, such as those blurred or covered by flies, were discarded. Finally, a total of 750 images were chosen.

All images were annotated using the popular annotation tool Labelme (<https://github.com/wkentaro/labelme.git>). On the one hand, the pixel-level cow masks were annotated with polygons, after which they were labelled either "Standing" or "Lying" according to their body postures. Cow instances were discarded for annotation when their posture was unable to recognise, e.g., being at the edge of the image or being too obscured by other cows or facilities. On the other hand, the flanks of lying cows were annotated as a new flank object detection dataset by using rectangles. When lying down, cows position their hind legs to one side, either stretched or not. Thus, the flank on the side where the hind legs were placed was labelled as "Flank_up" and the flank on the other side as "Flank_down". Some examples of annotation in both datasets are visualised in Fig. 1.

For both datasets, the annotated images were randomly split into training, validation, and testing sets at a ratio of 6:2:2. Training and validation sets were used to train the networks while the testing set was used to evaluate the final performance of the trained models. A detailed description of the datasets with regard to the number of instances or objects per class is given in Tables 1 and 2.

2.3. Multi-target RR measurement for lying cows based on deep learning and optical flow method

The overall working flow of the proposed computer vision-based RR measurement for multiple lying cows is shown in Fig. 2. All operations to be presented in this section were done in Python. Two programs were written, and ran in parallel, with the first one running two trained YOLO (You Only Look Once) models and the second one tracking predictions and calculating RR. Multithreading was used to speed up. The YOLO models were set to make inferences once every second to fulfil real-time measurements. All source code is available at <https://github.com/Kaiwen-Robotics/CattleRR.git>.



Fig. 1. Example annotations for cow instances (top) and flank objects (bottom). Polygons in red and green represent standing and lying cow instances, respectively. Rectangles in yellow and blue represent the flank objects on the hind leg side and the opposite side, respectively.

Table 1
Overview of the cow segmentation dataset.

Dataset	Number of cow instances		
	Standing	Lying	Total
Training (450 images)	1187	2476	3663
Validation (150 images)	411	808	1219
Testing (150 images)	407	788	1195
Total	2005	4072	6077

Table 2
Overview of the flank detection dataset.

Dataset	Number of flank objects		
	Flank-up	Flank-down	Total
Training (432 images)	1989	1417	3406
Validation (145 images)	693	501	1194
Testing (145 images)	734	508	1242
Total	3416	2426	5842

2.3.1. Architecture, training, and evaluation of deep learning networks

The YOLO series of algorithms is a popular framework for object detection that is well known for excellent performance and inference speed. YOLOv5, as a recent release, has gained significant attention in the computer vision community. Notably, its latest 7.0 version has been supported with instance segmentation and has been reported as a new state-of-the-art benchmark. Instance segmentation, as an extension of object detection, focuses not only on localising an object but also on generating a pixel-level mask for each object detected. As emphasised by the authors, these off-the-shelf models are roughly easy to train and deploy. Thus, YOLOv5 was used for object detection and instance segmentation tasks in this study. For clarity, instance segmentation networks were noted as YOLOv5-seg.

The training was performed in Python 3.7 language with Pytorch

1.7.1 on a computer with a 64-bit Windows 11 system and an NVIDIA GeForce RTX 3090 GPU. Transfer learning was used to accelerate the training process. The initialised weights used in this study were transferred from the network pre-trained on the Microsoft COCO benchmark. During the training process of the instance segmentation networks, a batch size of 4 was used, which was the maximum allowed by the hardware. In addition, a batch size of 16 was used for the object detection networks. In both cases, three sizes of the architecture were trained (i.e., YOLOv5s, YOLOv5m, YOLOv5l, YOLOv5s-seg, YOLOv5m-seg, and YOLOv5l-seg) and the epoch was always set to 100. Other hyperparameters were set as default.

The mean average precision (mAP), as the most commonly used performance metric for object detection and instance segmentation tasks, was used in this study. It builds on top of some basic concepts, including the Intersection of Union (IoU), precision (Pr), and recall (R). The IoU measures the intersection area between two bounding boxes or, two masks in the case of instance segmentation, as expressed in Eq. (2):

$$\text{IoU} = \frac{\text{Prediction} \cap \text{Groundtruth}}{\text{Prediction} \cup \text{Groundtruth}} \quad (2)$$

As a convention, an IoU threshold of 0.5 was used to qualify the prediction. For example, if the IoU between the ground-truth bounding box (or mask) and the predicted bounding box (or mask) was greater than 0.5, the predicted bounding box (or mask) was determined to be true positive (TP), otherwise it was false positive (FP). Additionally, a ground-truth bounding box (or mask) presented but not predicted was determined to be false negative (FN). Following that, Pr and R were determined, which represent how many positive predictions are actual positives and how many actual positives are correctly predicted as positive, respectively. Their equations are shown in Eqs. (3) and (4):

$$\text{Pr} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (3)$$

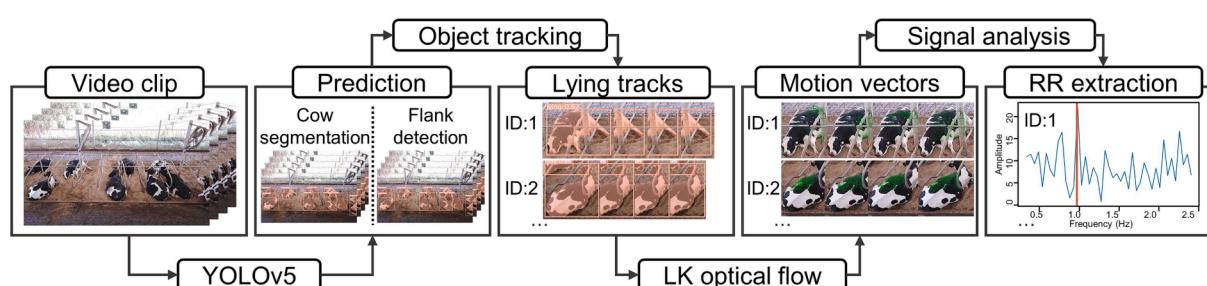


Fig. 2. Pipeline for multi-target respiration rate (RR) measurement for lying cows based on deep learning and optical flow method. Lucas-Kanade = LK.

$$R = \frac{TP}{TP + FN} \quad (4)$$

Subsequently, average precision (AP) was used to summarise Pr and R by computing the area under the Pr-R curve, and mean average precision (mAP) was used to summarise the AP over all classes. Their equations are shown in Eqs. (5) and (6):

$$AP = \int_0^1 Pr_{(R)} dR \quad (5)$$

$$mAP = \frac{1}{C} \sum_{i=1}^C AP_i \quad (6)$$

where C always takes 2, indicating 2 classes in both cases, i.e., “Lying” and “Standing” in the case of cow segmentation and “Flank_up” and “Flank_down” in the case of flank detection. The higher the mAP, the better the object detection or instance segmentation network. In addition, FPS was used to indicate the inference speed, as expressed in Eq. (7):

$$FPS = \frac{N}{t_N} \quad (7)$$

where t_N is the total inference time (s) on N images.

2.3.2. Optical flow

In this study, the optical flow method was used to track YOLO predictions, deal with camera shake, and extract respiration-related signals. Optical flow refers to the motion pattern of pixels in an image sequence resulting from either object or camera movement. It is represented as a two-dimensional vector field, with each vector representing the displacement of pixels from one frame to the next. Optical flow can be calculated using either sparse or dense methods. Sparse optical flow focuses on tracking a chosen set of pixels across frames, whereas dense optical flow aims to compute motion vectors for all pixels in the frame. The sparse optical flow was adopted in this study in consideration of efficiency.

Optical flow relies on several assumptions: (i) the pixel intensities of an object do not change between consecutive frames; (ii) the movement of the pixels over time is small.

Assumption (i) can be written as:

$$I(x, y, t) = I(x + dx, y + dy, t + dt) \quad (8)$$

where $I(x, y, t)$ represents the pixel intensity of a pixel with a coordinate of (x, y) at the frame t and it moves by (dx, dy) in the next frame after dt time.

According to assumption (ii), the image constraint at $I(x, y, t)$ to the right side of Eq. (8) can be expanded by the Taylor series as:

$$I(x + dx, y + dy, t + dt) = I(x, y, t) + \frac{\partial I}{\partial x} dx + \frac{\partial I}{\partial y} dy + \frac{\partial I}{\partial t} dt \quad (9)$$

that is to say, it follows:

$$\frac{\partial I}{\partial x} dx + \frac{\partial I}{\partial y} dy + \frac{\partial I}{\partial t} dt = 0 \quad (10)$$

divided by dt and results in:

$$\frac{\partial I}{\partial x} V_x + \frac{\partial I}{\partial y} V_y + \frac{\partial I}{\partial t} = 0 \quad (11)$$

where V_x and V_y represent the x and y components of the optical flow of $I(x, y, t)$, respectively, and $\partial I/\partial x$, $\partial I/\partial y$, and $\partial I/\partial t$ are the derivatives in the corresponding directions. Let $I_x = \partial I/\partial x$, $I_y = \partial I/\partial y$, and $I_t = \partial I/\partial t$, Eq. (11) can be further written as:

$$I_x V_x + I_y V_y + I_t = 0 \quad (12)$$

To solve the two unknowns (i.e., V_x and V_y), the Lucas-Kanade (LK) optical flow introduces another assumption that neighbouring pixels have similar motions. A search window is used to find neighbouring pixels around the target point, from which equations are then created, as shown in Eq. (13). Finally, V_x and V_y are solved by using the least squares principle. The LK optical flow was calculated by using the OpenCV function calcOpticalFlowPyrLK().

$$\begin{bmatrix} I_{x1} & I_{y1} \\ I_{x2} & I_{y2} \\ \vdots & \vdots \\ I_{xn} & I_{yn} \end{bmatrix} \begin{bmatrix} V_x \\ V_y \end{bmatrix} = \begin{bmatrix} -I_{t1} \\ -I_{t2} \\ \vdots \\ -I_{tm} \end{bmatrix} \quad (13)$$

2.3.3. Object tracking and stabilisation

Object tracking is necessary to link the predictions of each cow throughout the video clip. Tracking was done in a series of steps. First, the predictions of cow masks and their bounding boxes as well as those of the flank bounding boxes were obtained in each frame. To ensure reliable tracking, a minimum bounding box size threshold of 100 pixels was set to discard small objects. Next, the LK Optical Flow was used to track these detected objects across consecutive frames, mainly focusing on specific feature points of the objects identified by using the Shi-Tomasi corner detection method. At the same time, the flank mask was extracted by cropping the flank bounding box with the cow mask if the inclusion rate of the flank bounding box in the cow bounding box reached more than 90 %. The inclusion rate was calculated with Eq. (14):

$$Inclusionrate(ainb) = \frac{a \cap b}{a} \quad (14)$$

where a and b represent flank and cow bounding boxes, respectively. Tracks that were shorter than 10 frames were regarded as false positives and eliminated. Tracks with a missing flank mask for at least 10 frames were considered occluded and thus dropped. Tracks were split when labels were switched for at least 10 frames. Only tracks of “Lying” labels were entered into the following analysis.

Due to camera motion, there is some unwanted video shake existing in the current clips. Thus, an optical flow-based stabilisation was performed along with object tracking. First, the frames were initialised by converting to grayscale images, which simplified subsequent processing. Next, the Shi-Tomasi corner detection method was used to select feature points on the background area which was extracted by removing the cow mask from the bounding box, and the LK optical flows were calculated for the selected feature points. A transformation matrix was then computed based on the mean optical flows. Finally, the object was subjected to an affine transformation based on the translation matrix to obtain a stabilised alignment map as the output. In summary, the tracker tracked the motion of an object in a sequence of frames while the stabiliser aligned the object.

2.3.4. Optical flow calculation at abdominal areas

Previous studies have demonstrated that removing noise from the background and uninterested body areas increases the efficiency and accuracy of subsequent RR calculations (Guo et al., 2021). Wang et al. (2023) used oriented bounding boxes to localise animals in order to include animal pixels more accurately while excluding irrelevant background pixels. Our study adopted a different strategy to completely eliminate background noise by taking advantage of the instance segmentation technique. Specifically, the abdomen areas were obtained by intersecting the flank detections (from a YOLOv5 model) with the cow segmentations (from a YOLOv5-seg model) (Fig. 3).

Preliminary investigations show that the optical flow calculated by

using the LK sparse optical flow method specifically at the abdomen areas could better reflect respiration-related pixel movements (Fig. 3). LK sparse optical flow was used because it has shown better efficiency and accuracy in extracting respiration-related motions compared with dense methods (Song et al., 2019). Indeed, sparse optical flow is more computationally efficient, less susceptible to noise and occlusions, and requires less memory than dense optical flow.

Similarly to the processing of the stabiliser, the video frames were initialised by converting them to grayscale images. The difference is the location for calculating optical flows. Specifically, the gradient map of the grayscale images was first obtained and the background of the regions of interest (i.e., the intersections of cow masks and flanks) were excluded from the gradient map by setting them to zero. Next, grid points were obtained at a specified interval. LK sparse optical flow analysis was then performed between consecutive grayscale images using only grid points with a non-zero gradient.

2.3.5. Signal processing for extracting RR from the optical flow

The sinusoidal values of the resulting optical flow vectors were derived from the V_x and V_y . These sinusoidal values were then appended to a list that contained the sinusoidal values of all grid points across multiple frames. Subsequently, a Fast Fourier Transform was performed on each grid point's 120-frame sinusoidal values. A band-pass filtering of 0.4 to 2.4 Hz (i.e., 24 to 144 bpm) was applied to focus specifically on the respiration-related frequency range and remove unwanted noisy signals (e.g., occasional body wiggle and high-frequency abdominal twitching).

The frequency spectrums obtained from all grid points were then summed together. The instantaneous RR was determined by using peak detection on the summed spectrum, i.e., identifying the frequency with the maximum magnitude. This frequency was constantly updated and appended to a list of RR while processing the clip. The final RR across all frames was calculated by averaging the list of RR values at the end of the processing.

2.4. Performance analysis

A total of 55 video clips were randomly selected to measure RR by visual observation (Fig. 4). These clips had a mean \pm standard deviation duration of 16 ± 4 s and contained a total of 241 cow instances. First, lying cows were selected, that is 220 instances. Then, two trained observers with an intraclass correlation coefficient of 0.91 measured RR by manually counting the flank movement of the tracks over the clips and converting it to bpm. The results from the two observers were averaged. Cases with frequent and large unwanted movements, as well as those with visual occlusion of the abdomen, were discarded due to the inability to obtain ground-truth RR. As a result, ground-truth RR was obtained for a total of 180 tracks of lying cows. The mean \pm standard deviation number of available tracks per video clip was found to be 3.3 ± 1.7 , with a range between 1 and 8.

The statistical relationship between automated and manual RR measurements was evaluated through linear regression, with the Pearson correlation coefficient calculated using the cor function (R version 3.4.4; <https://R-project.org>). The agreement between automated and manual RR measurements was explored using a Bland-Altman plot and further quantified using the intraclass correlation coefficient calculated with the icc function from the "irr" package. Additionally, the overall inference speed was evaluated by using the average processing time and FPS of RR calculation on the 55 clips. Plus, an investigation was conducted to explore how the number of objects to be processed affected processing time and FPS.

In addition, a colour-based method was applied to extract respiration-related signals, as a comparison to the state of the art. The detailed methodology was adapted from Mantovani et al. (2023) which focused on non-contact RR measurement of group-housed lying cows. More specifically, the same object tracker and video stabiliser were used. The extraction of respiration-related signals was based on the averaged pixel intensity of the colour channels (R, G, B) of all pixels on the flank areas (i.e., intersection areas of flank detections and cow segmentations). RR was calculated using the same signal processing method.

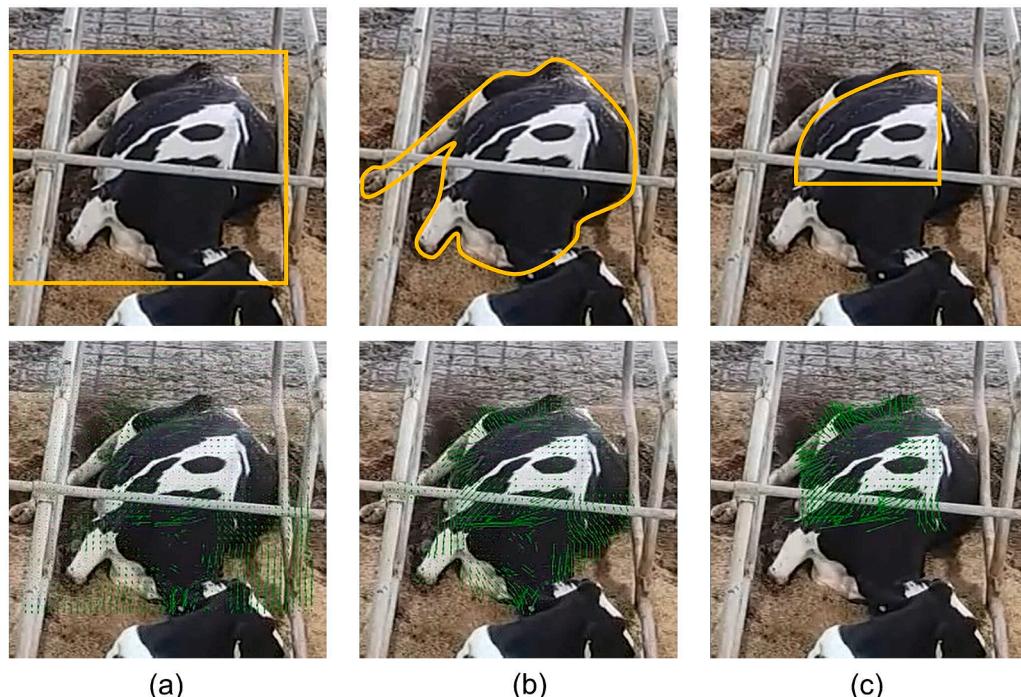


Fig. 3. Three regions of interest to compute sparse optical flows: (a) cow bounding box, (b) cow mask, and (c) cow mask overlapped by flank bounding boxes. The schematic of the regions of interest and the computed optical flows are shown in the top and bottom rows, respectively.

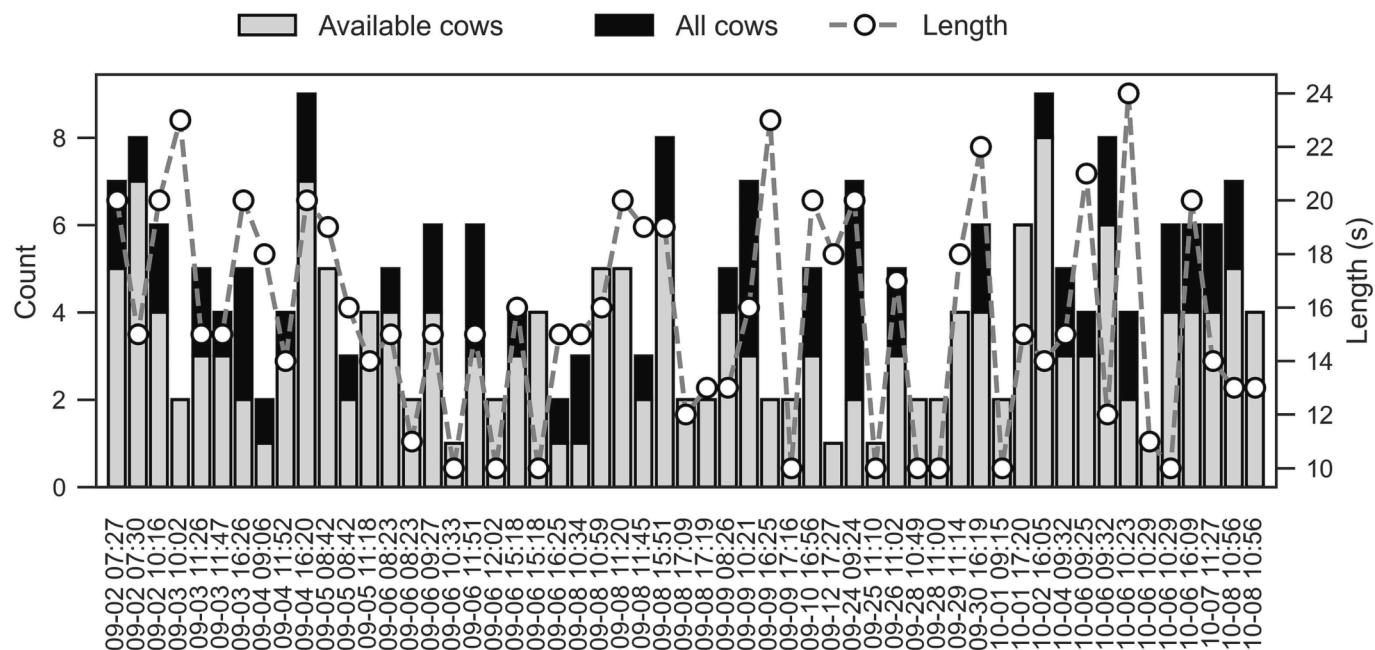


Fig. 4. Description of the dataset for the development of respiration rate measurement. The 55 video clips are sorted by datetime (mm-dd hh:mm).

2.5. Threshold development for heat stress

Piecewise regression models were used to fit the response of RR to heat stress (i.e., Ta and THI) and locate the breakpoint at which this response changed in trend. This was done by using the “segmented” package which determines the breakpoint based on the Davies test (Muggeo, 2008). Note that the analysis was based on the RR measured by the proposed framework on 55 video clips ($n = 180$). Specifically, simple linear regression models were first modelled to fit the response of RR to environmental variables (i.e., Ta and THI) using the lm function. Piecewise models were then built to update the simple models, written as follows:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 (X_i - X_{bp}) X_k + \varepsilon_i, X_k = \begin{cases} 0 & \text{if } X \leq X_{bp} \\ 1 & \text{if } X > X_{bp} \end{cases} \quad (15)$$

where Y is the RR, β_0 is the population intercept, X is the environmental variables (i.e., Ta and THI), X_{bp} is the breakpoint, X_k is the dummy variable, β_1 is the left slope, β_2 is the difference between right slope and left slope, and ε_i is the random residual for the i -th observation.

3. Results and discussion

3.1. Performance of cow segmentation and flank detection

As shown in Fig. 5, there was a rapid increase in validation mAP as training loss decreased for each model, and larger models always converged faster than smaller ones. The results of the testing sets demonstrate extremely good performance in segmenting cow instances with all three models achieving a bounding box mAP and a mask mAP exceeding 0.99 (Table 3). Additionally, all three object detection models had an mAP greater than 0.94 in detecting flank objects (Table 4). All these results once again confirm the strong capability of YOLOv5 to effectively learn and predict cattle-related datasets (Shu et al., 2023).

As expected, larger models had higher mAP but lower inference speed (Tables 3 and 4). In the case of cow instance segmentation, the YOLOv5s-seg model had a 0.1 % lower box mAP and 0.3 % lower mask mAP, but 120.8 % faster speed compared with the YOLOv5l-seg model. In the case of flank object detection, the YOLOv5s model had a 1.2 % lower mAP and 40.3 % faster speed compared with the YOLOv5l model.

Therefore, the smallest models in both cases (i.e., the YOLOv5s-seg and YOLOv5s models) were selected to be used for further RR measurement due to their sufficiently good performance and fastest speed. Some example predictions of the trained YOLOv5s-seg and YOLOv5s models on the testing set are shown in Fig. 6.

Many previous works have segmented pixel-level masks for cattle. For example, Salau and Krieter (2020) trained a Mask R-CNN network to segment Holstein-Friesian dairy cows housed in a similar free-stall barn. Their results showed a bounding box AP of 0.91 and a mask AP of 0.85. In addition, Xu et al. (2020) trained several Mask R-CNN networks to segment cattle in different outdoor scenarios. They reported a bounding box AP of 0.95 and a mask AP of 0.94 for cattle on pastures, and a bounding box AP of 0.91 and a mask AP of 0.90 for cattle in feedlots. However, it should be noted that the instance segmentation performance obtained by YOLOv5 in this study is much better than that obtained by Mask R-CNN in the abovementioned studies, suggesting YOLOv5 or its more recent release as a new benchmark for instance segmentation in cattle research.

3.2. Performance of RR measurement based on optical flow method

This study aimed to develop a computer vision-based framework for measuring RR in dairy cows lying in free stalls. As shown in Fig. 7 (a), the RR measured by the proposed optical flow-based method had a higher Pearson correlation coefficient ($r = 0.944$) and a lower root mean square error (5.35 bpm; RMSE) with the ground truth, outperforming the colour-based method. In addition, the proposed method had a narrower 95 % limits of agreement (-9.02 to 11.4 bpm) on the Bland-Altman plot with the differences distributed homogeneously along the mean difference-mean axis (Fig. 7 (b)), suggesting the absence of proportional error. The mean difference of 1.19 bpm means that the proposed method generally counted 1.19 more bpm than visual observation. Additionally, the proposed method's intraclass correlation coefficient with the visual observation (0.974) outperformed that of the colour-based method (0.869) by 10.5 %, confirming its superior agreement with the ground truth.

As for the overall inference speed, the optical flow-based method had an average processing time of 8.4 s and an average FPS of 64 per test video clip, which is 6.7 s and 26 FPS faster than the colour-based

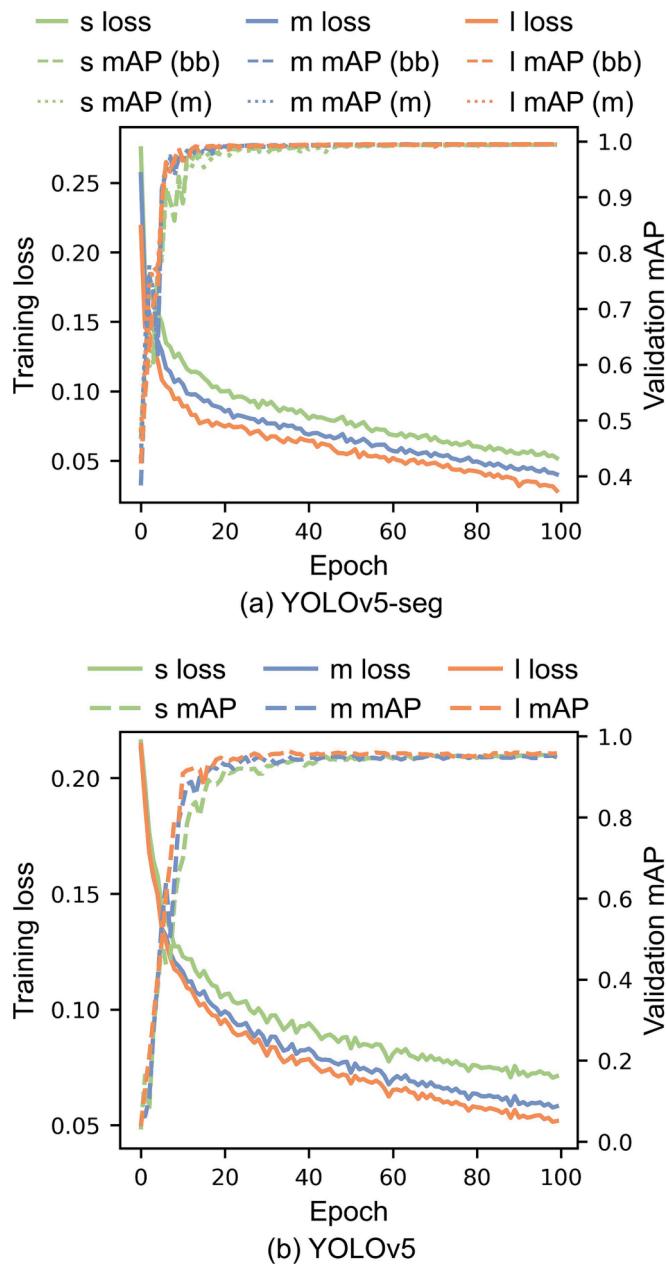


Fig. 5. Training loss and validation mean average precision (mAP) for (a) three YOLOv5-seg architectures (s, m, and l) for cow segmentation and (b) three YOLOv5 architectures (s, m, and l) for flank detection. bb in brackets represents bounding box and m in brackets represents mask.

method, respectively (Fig. 8 (a)). Both methods had a processing time positively related to the number of objects to be processed, with the proposed method being affected less, as indicated by the lower slope in Fig. 8 (b). Additionally, both methods had much lower processing FPS when handling a greater number of animals (Fig. 8 (c)). Anyway, all

clips processed with the optical flow-based method had an average FPS over 30, demonstrating the achievement of real-time measurement. As a control, 22 clips processed with the colour-based method had an average FPS below 30.

It should be acknowledged that previous studies have achieved optical flow-based RR measurement for an individual lying (Song et al., 2019) or standing (Wu et al., 2020) cow. However, these methods require the field of view to be filled with the region of interest (i.e., abdomen) which is less efficient for a large-scale dairy farm. Wu et al. (2023) achieved a multi-target RR measurement for cows in the exercise area by incorporating an instance segmentation network. However, their dataset only contained videos of two resting cows and the robustness of their method in free stalls and heat stress conditions remains unknown. Although our RMSE (5.35 bpm) is slightly higher than that of Wu et al. (2023) (3.74 bpm), our study covered much broader thermal conditions and achieved multi-object RR measurement for cows lying in an area of approximately 16 free stalls. It should also be noted that the abovementioned studies did not apply object tracking techniques since they used a pre-selected dataset with only fixed and limited objects. For such a real-world scenario in our study, we included an object tracker into the proposed framework to deal with the identity disorder due to changing positions of interfering objects, as well as a video stabiliser to deal with unwanted video shakes. In addition, the control method using the averaged pixel intensity of R, G, and B channels had a slightly higher RMSE of 9.56 bpm than 8.3 bpm in the original study of Mantovani et al. (2023), which may suggest that the current dataset is more challenging for colour-based methods to measure, possibly due to a larger distance between animals and the camera.

Theoretically, both optical flow- and colour-based methods may have issues where the estimated signal may not accurately represent body motions when being overexposed or having less body texture. However, the colour-based method was witnessed to suffer more since its calculation was based on the entire abdomen area which contained a bunch of pixels that did not change periodically with respiration. In contrast, optical flow-based methods pre-selected pixels in the abdomen area where the gradient was not zero. These pixels often belonged to the edges of patterns or dirt on the body, which had a higher amplitude of periodic changes, resulting in more stable RR estimation. In addition, the colour-based method's higher computational costs due to using all pixels for signal processing always led to lower processing speed. On the contrary, optical flow-based methods were faster since they only performed calculations on pixels with a non-zero gradient. It should be noted that our results do not necessarily indicate the superiority of

Table 4

Performance of the three YOLOv5 architectures for detecting flank objects on the testing set ($n = 145$).

Model	AP (bounding box)		mAP	FPS	Model size (MB)
	Flank-up	Flank-down			
YOLOv5s	0.967	0.924	0.945	108	13.6
YOLOv5m	0.971	0.942	0.956	85	40.1
YOLOv5l	0.978	0.936	0.957	77	99.4

AP = average precision; mAP = mean average precision; FPS = frames per second.

Table 3

Performance of the three YOLOv5 architectures for segmenting cow instances on the testing set ($n = 150$).

Model	AP (bounding box)		mAP	AP (mask)		mAP	FPS	Model size (MB)
	Standing	Lying		Standing	Lying			
YOLOv5s-seg	0.993	0.994	0.993	0.993	0.988	0.991	53	15.3
YOLOv5m-seg	0.993	0.995	0.994	0.993	0.995	0.994	36	42.6
YOLOv5l-seg	0.993	0.995	0.994	0.993	0.995	0.994	24	91.9

AP = average precision; mAP = mean average precision; FPS = frames per second.

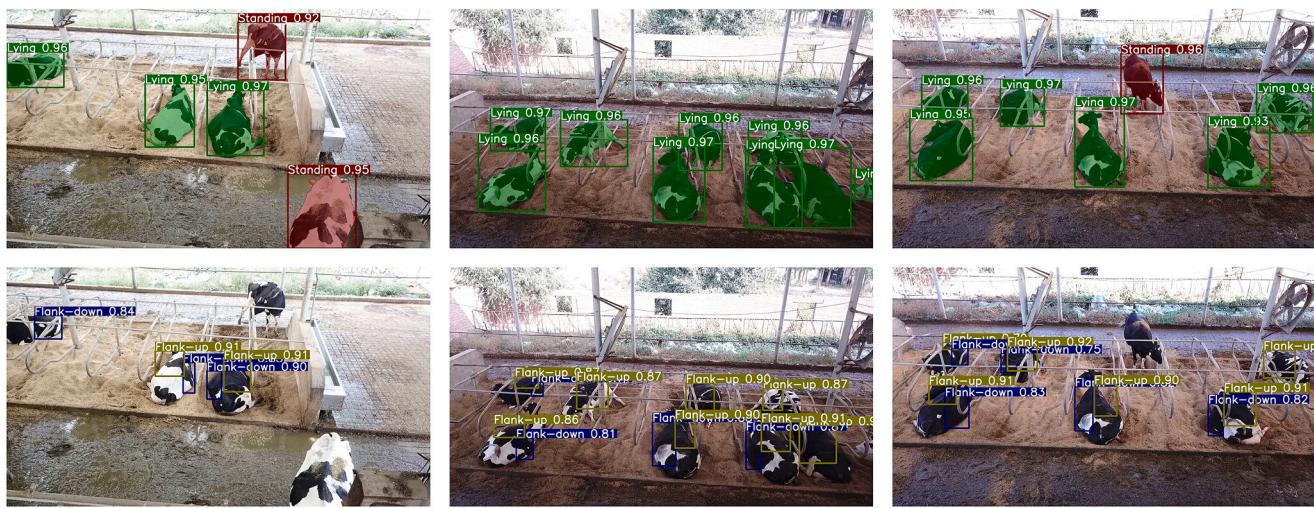


Fig. 6. Example predictions of cow segmentation (top) and flank detection (bottom) on the testing set. Predictions in red and green represent standing and lying cow instances, respectively. Predictions in yellow and blue represent the flank objects on the hind leg side and the opposite side, respectively.

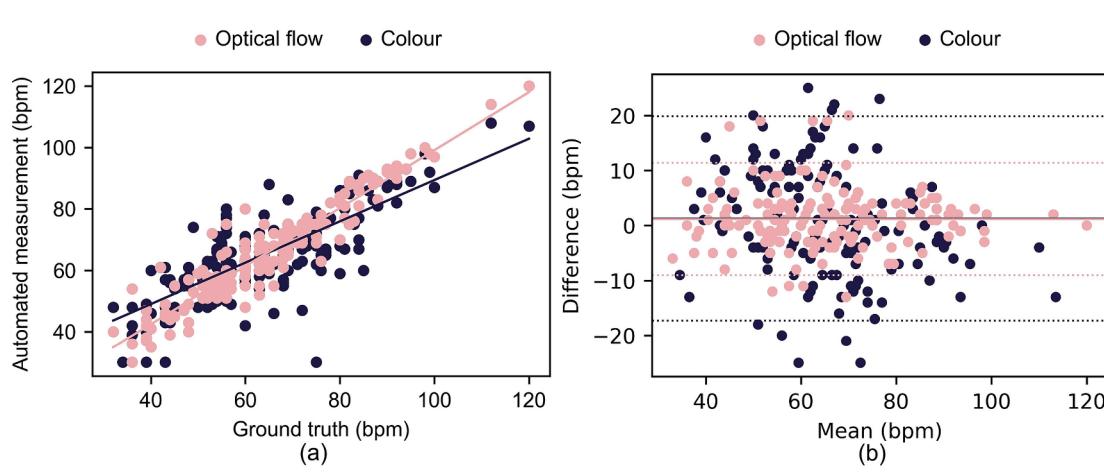


Fig. 7. Statistical analysis between automated and manual respiration rate measurements (breaths per minute (bpm), n = 180). Automated measurements relied on optical flow- and colour-based methods. (a) Linear regression plot, where the Pearson correlation coefficient of 0.944 and 0.796, respectively, and the root mean square error of 5.35 and 9.56 bpm, respectively. (b) Bland-Altman plot, where the solid lines represent the mean difference: 1.19 and 1.29 bpm, respectively, and the dashed lines represent 95 % limits of agreement: -9.02 to 11.4 bpm and -17.3 to 19.9 bpm, respectively.

optical flow methods over colour-based methods in all cases. It can only be concluded that colour-based methods require sophisticated filtering to select periodically changed pixels which can represent respiration. A possible solution, as suggested by Wang et al. (2023), is to filter the pixels by using standard deviation. However, it was not adopted as it failed to filter non-periodic noise in our case.

3.3. Response of RR to heat stress

The linear and piecewise regression models with Ta and THI as predictors are illustrated in Fig. 9 and detailed in Table 5. The slope of the model indicates a slow increase in RR of 0.32 bpm per unit increase in Ta below 23.6 °C and a much steeper increase in RR by 2.03 bpm per unit increase in Ta above 23.6 °C. This Ta threshold is slightly lower than the Ta threshold of 24.4 °C determined by Shu et al. (2022). It is probably because the current study included only lying cows which are more sensitive to heat stress (Pinto et al., 2020).

Additionally, RR increased slowly by 0.03 bpm per unit THI below 72 THI, and more steeply by 3.06 bpm per unit THI above 72 THI. This result is much higher than that of Pinto et al. (2020), who determined a

THI threshold of 65 for RR in lying cows. We speculate that the difference in the determined THI thresholds among studies can be attributed to the difference in production level since high productivity is well-known as a risk factor for heat stress (Collier et al., 2019). The cows in the four experimental pens had an average milk yield of 32.3 kg/d during the study, which is much lower than the herd in the study of Pinto et al. (2020) which had an average milk yield of 41.1 kg/d.

In addition, the determined THI threshold is lower than the threshold of 77 determined for dry cows (Ouellet et al., 2021). This is sensible since the lactating state of our experimental cows made them more sensitive to heat stress than dry cows. Other cow-related factors, such as body condition score, coat colour, and breed, can all have an effect on animals' heat tolerance (Gaughan and Castaneda, 2003). However, body condition and coat colour were not considered as grouping factors in this study, and no THI thresholds for RR for other breeds of dairy cows were found in previous studies. Thus, it is impossible to make a comparison with the current results.

The corresponding RR thresholds determined by the piecewise regression models were 60.9 and 60.2 bpm by Ta and THI, respectively (Table 5), which are consistent with Collier et al. (2012) who

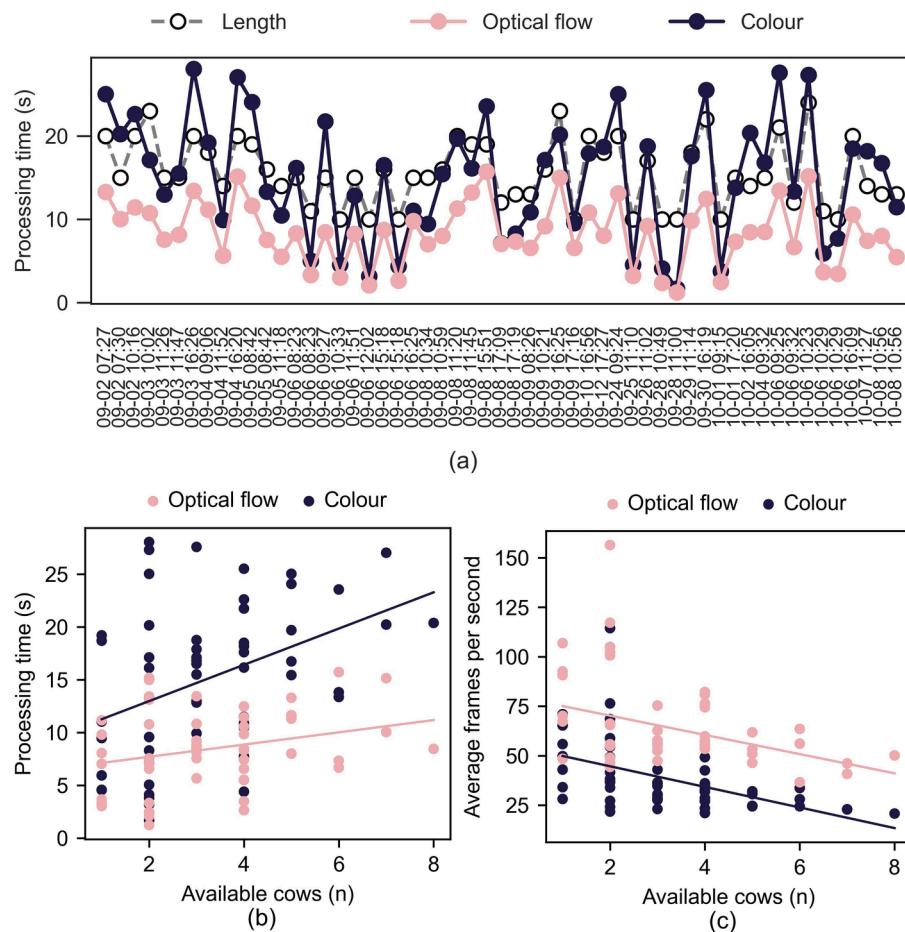


Fig. 8. Inference speed of the automated respiration rate measurements relied on optical flow- and colour-based methods. (a) Processing time of the 55 clips sorted by datetime (mm-dd hh:mm). (b) Relationship between the number of available cows to be processed and processing time. (c) Relationship between the number of available cows to be processed and average frames per second.

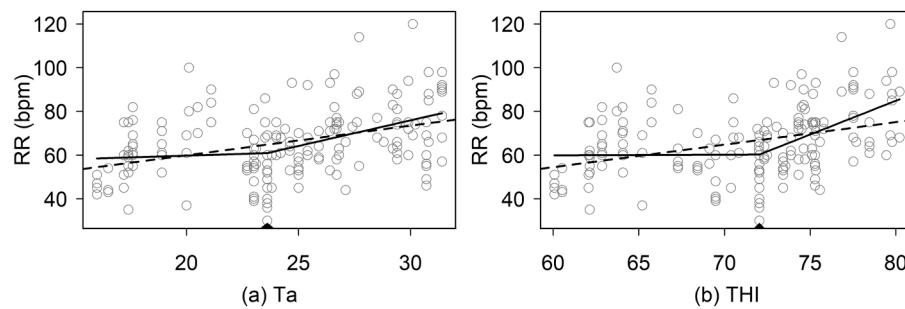


Fig. 9. Automated measurements of respiration rate (RR; breaths per minute (bpm), $n = 180$) and their fitted profiles from linear regression (dashed line) and piecewise regression (solid line) models with ambient temperature (T_a , °C) and temperature-humidity index (THI) as the predictors. The breakpoints are marked as black triangles above the x-axis.

determined the onset of heat stress when RR exceeded 60 bpm. In addition, the determined RR threshold is higher than those determined by Dalcin et al. (2016) for $\frac{1}{2}$ Holstein \times Girolando dairy cows (30) and $\frac{3}{4}$ Holstein \times Girolando dairy cows (45). This can be explained by the higher physiological states of pure-bred Holstein cows compared with mixed cows.

3.4. Limitations and perspectives

No sensitivity or comparison study was conducted to explore the impact of camera resolution and frame rate on the calculated RR. It is

also notable that high-end cameras are too expensive to provide a complete view of the barn. Network recording cameras, although more economically accessible in commercial farms, do suffer from practical issues such as more severe camera shake and lens distortion. In the next step, we will test and improve our method on a video recording system.

It should be noted that the proposed method was only trained and tested on the data collected from one free-stall barn with one specific herd and breed. Given the low generalisability of deep learning methods, further targeted training and testing are required before applying it to other facilities, herds, and breeds. Anyway, this can be accelerated by using transfer learning techniques.

Table 5

Parameter estimates (mean \pm SE) of the piecewise regression model predicting respiration rate with ambient temperature (Ta, °C) and temperature-humidity index (THI) as the predictor, respectively (n = 180).

Predictor	Intercept	Breakpoint	Left slope	ΔSlope	AIC	
					Linear	Piecewise
Ta	53.4 ± 11.3	23.6 ± 1.9	0.32 ± 0.56	2.03 ± 0.75	1466.2	1464.0
THI	58.0 ± 23.4	72.0 ± 1.2	0.03 ± 0.35	3.06 ± 0.75	1469.5	1455.8

ΔSlope = difference between right and left slopes; AIC = Akaike information criterion.

Additionally, night data was not collected in this study. Note that cows would take advantage of the greater temperature gradients from their bodies to the environment at night to relieve their accumulated heat load during the daytime (Allen et al., 2015). It is, therefore, necessary to complete 24-h measurements for continuous monitoring of cows' stress state. Although lights are turned on at night in the barn, the change in environmental illumination should have an impact on the robustness of the proposed method, which should be refined by further studies.

4. Conclusions

Collectively, our work represents an important step towards a non-contact RR measurement for dairy cows lying in free stalls. The proposed method can cover a lying area of approximately 16 free stalls and provide a precise and real-time RR measurement with a correlation coefficient of 0.944, an RMSE of 5.35 bpm, and an intraclass correlation coefficient of 0.974 when compared to visual observation, as well as an FPS of 64. Automated local cooling is promising by combining RR measurements and the critical thresholds to inform possible cooling systems at different places in the barn. However, further experiments and calibration with data collected by using low-cost video recording systems, recording during day- and night-time, are required before this technology can be applied on farms.

CRediT authorship contribution statement

Hang Shu: Writing – original draft, Software, Methodology, Investigation, Data curation, Conceptualization. **Jérôme Bindelle:** Conceptualization, Methodology, Writing – review & editing. **Xianhong Gu:** Writing – review & editing, Supervision, Resources, Funding acquisition, Conceptualization.

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.

Data availability

Data will be made available on request.

Acknowledgements

The first author (H. Shu) was funded by a PhD grant from China Scholarship Council [202203250080]. This work was also supported by the Beijing Dairy Industry Innovation Team Project [BAIC06-2016-2021] and the Agricultural Science and Technology Innovation Program [ASTIP-IAS07]. The authors are grateful to Dr. Gan Li, Dr. Alan, and

Kangyuan dairy farm for their assistance in data collection.

References

- Allen, J.D., Hall, L.W., Collier, R.J., Smith, J.F., 2015. Effect of core body temperature, time of day, and climate conditions on behavioral patterns of lactating dairy cows experiencing mild to moderate heat stress. *J. Dairy Sci.* 98, 118–127.
- Atkins, I.K., Cook, N.B., Mondaca, M.R., Choi, C.Y., 2018. Continuous respiration rate measurement of heat-stressed dairy cows and relation to environment, body temperature, and lying time. *Trans. ASABE* 61, 1475–1485.
- Bar, D., Kaim, M., Flamenbaum, I., Hanochi, B., Toaff-Rosenstein, R.L., 2019. Technical note: Accelerometer-based recording of heavy breathing in lactating and dry cows as an automated measure of heat load. *J. Dairy Sci.* 102, 3480–3486.
- Collier, R.J., Laun, W.H., Rungruang, S., Zimbleman, R.B., 2012. Quantifying Heat Stress and Its Impact on Metabolism and Performance. Florida Ruminant Nutrition Symposium. University of Florida, Gainesville, FL, USA, pp. 74–83.
- Collier, R.J., Baumgard, L.H., Zimbelman, R.B., Xiao, Y., 2019. Heat stress: physiology of acclimation and adaptation. *Anim. Front.* 9, 12–19.
- Dalcin, V.C., Fischer, V., Dalto, D.S., Alfonzo, E.P.M., Stumpf, M.T., Kolling, G.J., Silva, M.V.G.B.d., McManus, C., 2016. Physiological parameters for thermal stress in dairy cattle. *Revista brasileira de zootecnia* 45, 458–465.
- de Carvalho, G.A., Salman, A.K.D., da Cruz, P.G., de Souza, E.C., da Silva, F.R.F., Schmitt, E., 2020. Technical note: An acoustic method for assessing the respiration rate of free-grazing dairy cattle. *Livestock science* 241.
- Fuentes, S., Gonzalez Viejo, C., Tongson, E., Lipovetzky, N., Dunshea, F.R., 2021. Biometric Physiological Responses from Dairy Cows Measured by Visible Remote Sensing Are Good Predictors of Milk Productivity and Quality through Artificial Intelligence. *Sensors* 21, 6844.
- Gaughan, J., Castaneda, C., 2003. Refinement of Heat Load Index Based on Animal Factors. North Sydney NSW, Meat and Livestock Australia.
- Guo, T., Lin, Q., Allebach, J., 2021. Remote estimation of respiration rate by optical flow using convolutional neural networks. *Electronic Imaging* 2021, 267–261–267–211.
- Jorquera-Chavez, M., Fuentes, S., Dunshea, F.R., Warner, R.D., Poblete, T., Morrison, R.S., Jongman, E.C., 2020. Remotely Sensed Imagery for Early Detection of Respiratory Disease in Pigs: A Pilot Study. *Animals* 10, 451.
- Lowe, G., Sutherland, M., Waas, J., Schaefer, A., Cox, N., Stewart, M., 2019. Infrared Thermography-A Non-Invasive Method of Measuring Respiration Rate in Calves. *Animals* 9.
- Mantovani, R.R., Menezes, G.L., Dórea, J.R.R., 2023. Predicting respiration rate in unrestrained dairy cows using image analysis and Fast Fourier Transform. *JDS Communications*.
- Milan, H.F.M., Maia, A.S.C., Gebremedhin, K.G., 2016. Technical note: Device for measuring respiration rate of cattle under field conditions 1. *J. Anim. Sci.* 94, 5434–5438.
- Muggeo, V.M., 2008. Segmented: an R package to fit regression models with broken-line relationships. *R News* 8, 20–25.
- Nrc, 1971. A Guide to Environmental Research on Animals. National Academy Press, Washington, DC, USA, p. 374.
- Ouellet, V., Toledo, I.M., Dado-Senn, B., Dahl, G.E., Laporta, J., 2021. Critical Temperature-Humidity Index Thresholds for Dry Cows in a Subtropical Climate. *Frontiers in Animal Science* 2.
- Pastell, M., Aisla, A.M., Hautala, M., Poikalainen, V., Praks, J., Veermäe, I., Ahokas, J., 2006. Contactless measurement of cow behavior in a milking robot. *Behav. Res. Methods* 38, 479–486.
- Pinto, S., Hoffmann, G., Ammon, C., Amon, T., 2020. Critical THI thresholds based on the physiological parameters of lactating dairy cows. *J. Therm. Biol.* 88.
- Salau, J., Krieter, J., 2020. Instance Segmentation with Mask R-CNN Applied to Loose-Housed Dairy Cows in a Multi-Camera Setting. *Animals* 10, 2402.
- Shu, H., Guo, L., Bindelle, J., Fang, T., Xing, M., Sun, F., Chen, X., Zhang, W., Wang, W., 2022. Evaluation of environmental and physiological indicators in lactating dairy cows exposed to heat stress. *Int. J. Biometeorol.* 66, 1219–1232.
- Shu, H., Bindelle, J., Guo, L., Gu, X., 2023. Determining the onset of heat stress in a dairy herd based on automated behaviour recognition. *Biosys. Eng.* 226, 238–251.
- Song, H., Wu, D., Yin, X., Jiang, B., He, D., 2019. Respiratory behavior detection of cow based on Lucas-Kanade sparse optical flow algorithm. *Transactions of the Chinese Society of Agricultural Engineering* 35, 215–224.
- Strutzke, S., Fiske, D., Hoffmann, G., Ammon, C., Heuwieser, W., Amon, T., 2019. Technical note: Development of a noninvasive respiration rate sensor for cattle. *J. Dairy Sci.* 102, 690–695.
- Tuan, S.-A., Rustia, D.J.A., Hsu, J.-T., Lin, T.-T., 2022. Frequency modulated continuous wave radar-based system for monitoring dairy cow respiration rate. *Comput. Electron. Agric.* 196, 106913.
- Wang, M., Li, X., Larsen, M.L.V., Liu, D., Rault, J.-L., Norton, T., 2023. A computer vision-based approach for respiration rate monitoring of group housed pigs. *Comput. Electron. Agric.* 210, 107899.
- Wu, D., Yin, X., Jiang, B., Jiang, M., Li, Z., Song, H., 2020. Detection of the respiratory rate of standing cows by combining the Deeplab V3+ semantic segmentation model with the phase-based video magnification algorithm. *Biosys. Eng.* 192, 72–89.
- Wu, D., Han, M., Song, H., Song, L., Duan, Y., 2023. Monitoring the respiratory behavior of multiple cows based on computer vision and deep learning. *J. Dairy Sci.* 106, 2963–2979.
- Xu, B., Wang, W., Falzon, G., Kwan, P., Guo, L., Chen, G., Tait, A., Schneider, D., 2020. Automated cattle counting using Mask R-CNN in quadcopter vision system. *Comput. Electron. Agric.* 171, 105300.