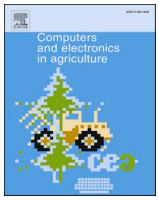
Computers and Electronics in Agriculture 192 (2022) 106517

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag



Comparison of the automated monitoring of the sow activity in farrowing pens using video and accelerometer data

Maciej Oczak [a,b,*], Florian Bayer [b], Sebastian Vetter [b], Kristina Maschat [b,c], Johannes Baumgartner [b]

[a] Precision Livestock Farming Hub, The University of Veterinary Medicine Vienna (Vetmeduni Vienna), Veterinärplatz 1, 1210 Vienna, Austria
[b] Institute of Animal Welfare Science, The University of Veterinary Medicine Vienna (Vetmeduni Vienna), Veterinärplatz 1, 1210 Vienna, Austria
[c] Austrian Competence Centre for Feed and Food Quality, Safety and Innovation, FFoQSI GmbH, Technopark 1C, 3430 Tulln, Austria

ARTICLE INFO

Keywords:
Sow
Activity
Computer vision
Accelerometer
Automated monitoring
Deep learning

ABSTRACT

Patterns in pigs activity can be an indicator of health and welfare of the animals. This motivates researchers to develop Precision Livestock Farming (PLF) tools for automated monitoring of pig activity level. In this research we compared two important technologies that can be used for this purpose, ear tag accelerometer and computer vision. Additionally, we compared both technologies with gold standard based on human labelling. A state-of-the-art object detection algorithm RetinaNet was trained on 9969 images and validated on 4273 images to automatically detect head of a sow, body of a sow, left ear, right ear and a hay rack. It was possible to detect these objects with a performance of 0.26 mAP@0.5:0.95. Activity of 6 sows was derived from detected parts of animals' bodies and compared with activity measurement based on ear tag accelerometer data. Dynamic relation between activity measurement based on both technologies was modelled with Transfer Function (TF) models. For all 6 animals activity of the body of a sow based on object detection was very similar to accelerometer based activity measurement ($R^2 > 0.7$). Similarly R^2 between activity of a head of a sow and accelerometer based activity was also very similar for most sows ($R^2 > 0.7$). Results of fitting of TF models to animal activity data based on ear tag accelerometer and output of object detection on body of sows and head of sows suggests that both technologies, the accelerometer and computer vision provide very similar information on activity level of animals. The presented computer vision method is limited to monitoring one animal under camera view as detected body parts cannot be associated with multiple individuals. Moreover, we expect that the method requires re-training the RetinaNet object detection algorithm with additional images collected on additional farms to achieve satisfactory performance in different environments. Application of computer vision approach might be advantageous in some PLF applications as it is non-invasive and might be less laborious than method based on ear tag accelerometer data.

1. Introduction

Patterns in activity of pigs such as reduced or increased activity level over a period of several days or changes in average daily activity levels are related to health and welfare of the animals and can be an indicator of sickness (Hart, 1988). When animals are sick they will reduce their overall physical activity, such as locomotion, sexual behavior, exploration, aggression, food and water intake, and social behaviour (Lopes, 2017). By adopting this strategy an animal conserves energy and consequently improves its ability to fight an infection (Hart, 1988). Sick animals divert resources to those functions of critical short-term value as e.g. maintaining body temperature and reducing behavioural traits that offer only longer-term fitness and may be considered less urgent at least under some conditions (Weary et al., 2009). Decreased activity levels in pigs can indicate lameness (Contreras-Aguilar et al., 2019). Reduction of locomotion ability in relation to lameness in pigs is directly caused by pain or general discomfort (Heinonen et al., 2013) but can also be interpreted as part of an animal's strategy to conserve the energy, protect the limbs and improve chances for recovery. Recovered animals increase their activity to the same level as non-lame pigs (Ala-Kurikka

* Corresponding author at: Precision Livestock Farming Hub, The University of Veterinary Medicine Vienna (Vetmeduni Vienna), Veterinärplatz 1, 1210 Vienna, Austria.
E-mail address: Maciej.Oczak@vetmeduni.ac.at (M. Oczak).

https://doi.org/10.1016/j.compag.2021.106517
Received 23 July 2021; Received in revised form 22 September 2021; Accepted 24 October 2021
Available online 2 December 2021

et al., 2017).

Similarly pigs that recover from heat stress increase their activity levels to normal (Von Jasmund et al., 2020). Pigs are relatively sensitive to high environmental temperatures because they cannot sweat. Lack of proper thermoregulation leads to heat stress and suffering of the animals (Huynh et al., 2005). To decrease production of heat and improve thermoregulation pigs reduce activity during heat stress.

Before the beginning of farrowing sows perform nest-building behaviour, which is mainly motivated by newborn piglets' need for protection against unfavourable weather conditions, predators and trampling by other adult pigs in natural conditions (Jensen, 1986). Nest-building behaviour occurs due to hormonal changes and the presence of external stimuli and results in increased activity level (Castrén et al., 1993). Activity level might be more than two times higher on the day of farrowing, when sows build nests the most intensely, in comparison to earlier days before farrowing (Oliviero et al., 2008).

Above-named examples of relations between health and welfare of pigs and patterns in their activity motivate researchers in the public and private sector (Smartbow GmbH, Weibern, Austria; Ro-Main, Saint-Lambert-de-Lauzon, Canada) to develop Precision Livestock Farming (PLF) technologies (Wathes et al., 2008) for automated monitoring of pig activity level. Sensor technologies have enormous potential to support farmers, veterinarians or the other stakeholders in the early detection of health and welfare problems in livestock. PLF technologies are able to detect subtle changes in behaviour and pre-pathological indicators of sickness days or weeks before they become apparent to humans (Martínez-Avilés et al., 2017). Furthermore, early detection of behavioural changes can not only support disease prevention but also hint at management and husbandry drawbacks that might lead to serious welfare concerns (Chapa et al., 2020). Specific sensors used for monitoring pig activity are infrared photocells (Erez and Hartsock, 1990), force sensors (Oliviero et al., 2008), passive infrared sensors (PIR) (Von Jasmund et al., 2020), tri-axial accelerometers (Cornou and Lundbye-Christensen, 2008) and 2D cameras (Costa et al., 2009).

In this research we focus on comparison of automated measurement of activity level of sows in the pre-farrowing period, with two technologies. We compare measurement of sow activity in the farrowing environment based on tri-axial accelerometer sensors mounted in the ear tags with the computer vision approach. We focus on tri-axial accelerometers mounted in an ear tag because position of the sensor in any other way than in ear tags, e.g. in collars or on the legs of pigs, creates a danger of damage or reposition of accelerometers by the other pigs (Hamäläinen et al., 2011). Thus, for the practical application of accelerometer based techniques, for monitoring activity level, ear tag might be the most desirable solution (Cornou and Lundbye-Christensen, 2008).

The potential advantage of computer vision approach for measurement of activity of pigs in comparison to tri-axial accelerometer, ear tag based approach is that it allows monitoring in a non-invasive way, without human presence or with limited human presence. It is possible to monitor multiple animals with one sensor (Liu et al., 2020). Camera sensor and computer vision is also probably the most versatile monitoring technology applied so far in PLF. Thus, meaning that multiple variables (key indicators) related to health and welfare can be monitored with one camera sensor (Chen et al., 2020). The most important potential advantage of computer vision approach for measurement of activity of pigs is that computer vision approach allows to directly measure the movement of different body parts of a pig such as head, ears or center of the body, while accelerometer placed in an ear tag allows measurement of movement of an ear of the animal and based on this measurement it is possible to derive how active the animal is.

The first objective of this study was to apply a state-of-the-art object detection algorithm for measurement of activity of sows in the farrowing pens, in the pre-farrowing period. Object detection should allow measurement of activity of sows by monitoring movement of their different body parts such as head, center of the body or ears. The second objective was to compare the similarity of measurement of activity of sows between ear tag, tri-axial accelerometer, gold standard based on human observation and automated camera based method.

2. Materials and methods

2.1. Ethical statement

Project PIGwatch was authorized by the Ethical Committee of the Austrian Federal Ministry of Science, Research and Economy and by the Ethical Committee of Vetmeduni Vienna (GZ: BMWFV-68.205/0082-WF/II/3b/2014) according to the Austrian Tierversuchsgesetz 2012, BGBl. I Nr. 114/2012.

2.2. Experimental setup

2.2.1. Animals and housing

The experiment was conducted between June 2014 and May 2016 at the pig research and teaching farm (VetFarm) of the University of Veterinary Medicine Vienna, Vienna, Austria. In total, 36 Austrian Large White sows and Landrace × Large White crossbreds sows from five days before farrowing to the end of the four weeks lactation period were included in the experiment. These sows were housed in three types of farrowing pens, which offered the option of either keeping the sows free or in a farrowing crate. Out of 33 sows, 11 were kept in SWAP (Sow Welfare and Piglet Protection) pens (Jyden Bur A/S, Vemb, Denmark), 11 in trapezoid pens (Schauer Agrotronic GmbH, Prambachkirchen, Austria) and 11 in wing pens (Stewa Steinhuber GmbH, Sattledt, Austria). None of the animals included in the experiment were confined in a farrowing crate from the introduction to the farrowing pen until the end of farrowing. Sows were housed in three types of farrowing pens to increase variability of housing conditions. This should increase robustness of any algorithms developed on collected dataset.

The SWAP pens had an area of 6.0 m2. The pens had a solid concrete floor in the front (lying area) and a slatted cast iron floor in the back (defecation area). The pen had 2 troughs, one for the crated and one non-crated sow (Fig. 1a). The trapezoid pens had an area of 5.5 m2. The pens had plastic flooring in the creep area and solid concrete flooring in the sow lying area in front of the trough (Fig. 1b). The wing pens had an area of 5.5 m2. The pens were partly slatted with plastic elements and solid concrete elements (Fig. 1c). In all 3 pen types a straw rack was mounted in the front area of the pen, in close proximity to the trough.

The sows were introduced to the farrowing pens approximately five days before the expected date of farrowing. The date was derived from the usual gestation length of sows (114 days), which could vary from 105 to 125 days. The experimental period was from the introduction of the sow to the farrowing room until 24 h after the end of farrowing. The experimental pens were located in a testing unit of the VetFarm which had an automatic ventilation system. The average temperature in the room was 22 °C. The sows were fed twice a day during the experimental period. Water was provided permanently in the troughs via a nipple drinker or an automatic water-level system. To fulfil the need for adequate material to explore and for nest building, sows and piglets were offered straw in the aforementioned rack throughout their stay in the pens (Fig. 2). The racks were half filled in the morning and whenever the racks were empty.

2.2.2. Video recording

Behaviour of sows was video recorded from introduction to the farrowing pens until weaning with 2D cameras in order to create a data set that could be labelled. Each pen was equipped with one IP camera (GV-BX 1300-KV, Geovision, Taipei, Taiwan) locked in protective housing (HEB32K1, Videotec, Schio, Italy) hanging 3 m above the pen, giving an overhead view. Additionally, infrared spotlights (IR-LED294S-90, Microlight, Bad Nauheim, Germany) were installed in order to allow night recording. The videos were recorded with 1280x720 pixel resolution, in MPEG-4 format, at 30 fps.

Fig. 1. Farrowing pens with possibility of temporary crating. (a) SWAP pen, (b) trapezoid pen, and (c) wing pen.

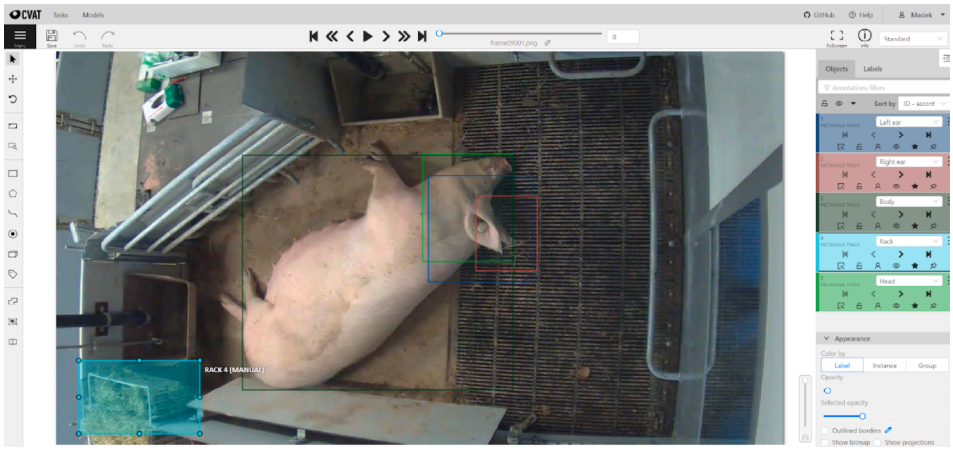


Fig. 2. Labelling of a video frame in CVAT. Objects labelled: left ear, right ear, head, body and the hay rack. A hay rack is highlighted in the figure as it is selected by the mouse pointer. Ear tag with an accelerometer is placed on the left ear of the sow.

The cameras were connected to a PC on which Multicam Surveillance System (8.5.6.0, Geovision, Taipei, Taiwan) was installed. The system allowed simultaneous recording of videos from 9 cameras. The PC had a Intel i5-3330 processor, 3 GHz (Intel, Santa Clara, United States of America) with 4 GB of physical memory. The operating system was Microsoft Windows 7 Professional (Redmond, United States of America). Recordings were stored on exchangeable, external 2 and 3 TB hard drives.

2.2.3. Accelerometer system

The SMARTBOW® system (Smartbow GmbH, Weibern, Austria) consisting of ear tags, wall points, and station delivered information on acceleration of sensors attached to the sows. The SMARTBOW® ear tag was equipped with an accelerometer sensor that measures acceleration in 3 axes (xyz). The accelerometer can record acceleration with ±2 g (1 gravity of Earth = 9.80665 m/s^2) There were five SMARTBOW® antennas installed on the walls of the experimental compartment. Data loss was not >10% in the experimental period.

2.3. Dataset

The dataset composed of video of 33 sows recorded in a period from introduction to farrowing pen until 24 h after the end of farrowing was divided into a subset for training and validation of object detection algorithm and a subset for comparison of measurement of activity level by computer vision with measurement based on ear tag accelerometer and gold standard based on human observation. The subset for training and validation of object detection algorithm consisted of 27 (82%) animals, while the subset for comparison of measurement of activity level by computer vision with measurement based on ear tag accelerometer and gold standard consisted of 6 (18%) animals (Table 1). We divided the dataset into these two subsets to make sure that the part of the whole dataset used for comparison of measurement of activity level by computer vision with measurement based on ear tag accelerometer and gold standard was not used to train and validate the object detection algorithm. This division of the dataset increases the validity of conclusions drawn from our research as both data subsets were independent.

Table 1
Dataset divided into subset for training and validation of object detection algorithm and subset for comparison of computer vision and ear-tag accelerometer data.

Pen Type	Training and validation of object detection algorithm	Comparison of computer vision and ear-tag accelerometer
SWAP	9	2
Trapezoid	9	2
Wing	9	2
Total	27	6

The animals in both subsets were equally distributed between SWAP, trapezoid and wing pens. (Table 1). The final dataset gave a possibility to compare both the measurement of activity data based on computer vision with ear tag accelerometer data and gold standard from introduction into the farrowing pen until 24 h after the end of farrowing.

2.4. Data labelling

Video recorded sow behaviour was manually labeled in order to create a reference dataset on the basis of which further data analysis could be performed. In the first step of the labeling process, the time of the onset of farrowing of each individual sow (n = 33) was labeled. The onset of farrowing was defined as the point in time when the body of the first piglet born dropped on the floor. The time of birth of the last piglet indicated the end of farrowing. Gold standard for estimation of activity

level of sows was based on manual labelling of standing and walking behaviours by trained labeller. Gold standard was labeled for all sows (n = 6) included in the subset for comparison of measurement of activity level by computer vision with measurement based on ear tag accelerometer. Walking behaviour was defined as movement for at least 2 s forward, backward or to the side. Stepping from one foot to the other was not included. (modified after Johnson et al., 2007). Start: a sow lifts her first leg. End: all of sows feet are back on the ground again. Standing behaviour was defined as all four limbs of the sow are in contact with the floor or the sow is stepping from one foot to the other so that no forward or backward or side movement can be seen. Her body is lifted up in the air (modified after Baxter et al., 2011). Start: all four extremities have contact with the floor. End: The weight of the body rests on the hind limbs, the hind touches the floor. Standing and walking behaviours had to last at least 2 s to be recorded by a labeller.

Labeling software Interact (version 9 and 14, Mangold International GmbH, Arnstorf, Germany) was used to label the beginning and end of farrowing on recorded videos. Labeling software Boris (version 7.9.15, Torino, Italy) was used to label gold standard for estimation of activity level of sows.

2.4.1. Frame selection for labelling

To train an object detection algorithm for automated detection of parts of sow's body it was necessary to select the least number of frames in the dataset containing the most relevant information on the objects of interest. This general approach allows the most efficient use of computational resources for training of the object detection algorithm but also reduces the heavy workload related to manual labelling of objects in the images by human labellers.

In the first step we selected video data recorded for 3 days for each animal in the subset for training and validation of object detection algorithm (26 out of 33 sows). The 1st day out of 3 selected days was the day of introduction to the farrowing pen, the 2nd selected day was the day of farrowing and the 3rd day was one day before the day of farrowing. The duration of the first period was around 12 h and the duration of the two remaining periods was 24 h each. The first day was of around 12 h duration as sows were moved into the farrowing compartment in the middle of the day, leaving around 12 h until the midnight of that day.

As pens become less clean, objects in a pen change their appearance which might affect the performance of object detection algorithm. Selection of these 3 specific periods allowed us to represent variable cleanliness of the pen, with a very clean pen on the day of the introduction of a sow to the pen and slowly degrading cleanliness on the following days.

We also assumed that piglets might look similar to the parts of the body of a sow such as ears or legs. Thus, including the day of farrowing in the dataset allowed the object detection algorithm learning to differ between the piglets and sow body parts. This could prevent some of the possible misclassification between sow body parts and the piglets if the object detection algorithm were to be applied with piglets present in the pen. Finally, we included the period of one day before the day of farrowing as on that day sows are the most active and express the highest variability of postures and behaviours due to nest-building activity (Castrén et al., 1993). We expected that increasing the variability in postures and behaviours of sows in the dataset would increase the robustness of the object detection algorithm.

Labelling 100 frames with 5 selected object classes (left ear, right ear, head, body and a rack) took around 1 h for one human labeller (Fig. 2). For the purpose of selection of specific frames to be used for labelling we applied the k-means algorithm described in Pereira et al. (2019). K-means algorithm was used to reduce correlation between sampled images. Out of around 175 000 000 frames (24 h + 24 h + 12 h = 60 h * 60 = 3600 min * 60 = 216 000 s * 30 frames/s = 6 480 000 * 27 sows = 174 960 000 frames) we decided to select 14 242 frames as it was possible to label this number of images within 3.5 weeks by one labeller (40 h/week) and that was manageable with the resources available in this research project. The k-means algorithm identified 14 242 frames that were the most different between each other in our dataset. These frames were later used for labelling by human labeller and also for training and validation of the RetinaNet object detection algorithm.

2.4.2. Labelling of a single frame

Five object classes were labelled by a trained human labeller on each frame out of selected 14,242 frames, left ear, right ear, head and body of the sows and the hay rack in the pen (Fig. 2). Computer Vision Annotation Tool (CVAT) was used to label the frames. Each of the 5 object classes was labelled with a rectangle so that the center of an object was placed in the center of the rectangle. Left and right ears were labelled with an area surrounding them. The size of that area had the width of the labelled ear on the image. This was done as the performance of the object detection algorithm was better when a bit larger area around the ears was labelled. The other objects, head, body and the hay rack were labelled so that the boundaries of the rectangles fitted exactly with the boundaries of labelled objects (Fig. 2).

We decided to include the ears and the head of the sow in the labelling process assuming that movement of these parts of the body should be the most similar to the movement recorded by the accelerometer, which was placed on one of the ears of a sow.

To verify if the computer vision approach can provide additional information on animal activity we decided to also include the whole body of a sow in the labelling process. Movement of the body of a sow might be related mainly to the change of location of a sow in the pen. Thus, it was interesting for us how similar activity patterns measured on the body of the sow were in comparison to the measurements based on the head, ears and by ear tag accelerometer. Finally, the hay rack used for providing a sow with nest building material was labelled to compare performance of object detection algorithm between moving objects such as head or ears and a stationary object.

2.5. Object detection algorithm

We decided to apply Pytorch implementation of RetinaNet object detection algorithm (source code available at https://github.com/yhenon/pytorch-retinanet) for the task of detecting parts of the body of a sow and the hay rack in a farrowing pen. RetinaNet was first proposed by Lin et al. (2017) as a one-stage object detector (Fig. 3). The central reason for lower accuracy of one-stage detectors, the extreme foreground-background class imbalance, was addressed in RetinaNet by reshaping the standard cross entropy loss such that it down-weights the loss assigned to well-classified examples. RetinaNet is able to match the speed of previous one-stage detectors while surpassing the accuracy of all existing state-of-the-art two-stage detectors (Lin et al., 2017). This makes this object detection algorithm potentially usable in PLF applications in which real-time processing is of high importance.

Our labelled dataset with 14,242 images was divided into a training set with 9969 (70%) images and a validation set with 4273 (30%) images. Images were assigned randomly to the training and the validation sets. To compare the performance of RetinaNet on variable number of training images we trained the object detector on an increasing number of frames, 39, 78, 156, 312, 623, 1246, 2492, 4984 and 9969. Evaluation of performance of object detection algorithm on a variable number of training images allows estimation of effort needed for labelling of images in relation to expected performance of the algorithm. Each of nine trained models was evaluated by standard Common Objects in COntext (COCO) metric mAP@0.5:0.95, which averages average precision (mAP) over Intersection over Union (IoU) from 0.5 to 0.95 (Lin et al., 2014). With this metric we verified how well the estimated algorithm could detect objects on the training set and the validation set. MAP@0.5:0.95 represents the performance of the object detector for all 5 labelled object classes, left ear, right ear, head, body and the rack. Beside mAP@0.5:0.95 the average precision (AP@0.5:0.95) was

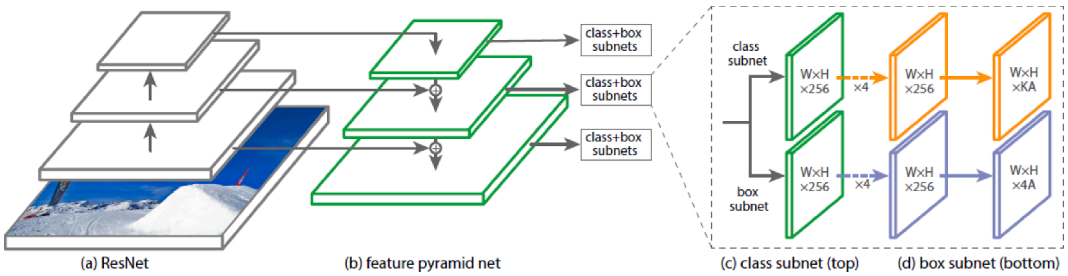


Fig. 3. The one-stage RetinaNet network architecture, from Lin et al. (2017).

calculated separately for the final models for each object class.

Each RetinaNet model was trained for 90 epochs with a standard learning rate of 0.00001. A pre-trained ResNet-152 backbone (ImageNet database) was used (Russakovsky et al., 2015). Images were normalized and augmented by flipping images horizontally with probability of 50% before processing. Adam optimization algorithm was used. ResNet-152 was used as it is the deepest network of the ResNet family and potentially can achieve better mAP in comparison to the other ResNet networks (He et al., 2016).

RetinaNet models were trained on a workstation with two CPUs Intel Xeon Gold 6226 (total 24 cores), 256 GB memory and a NVIDIA Titan RTX GPU 24 GB GDDR6. It was possible to train the models with 3.5 fps on the Titan RTX GPU installed on the workstation. Thus, training of the RetinaNet model on 9969 frames for 90 epochs took approximately 3 days.

2.6. Activity based on object detection

To extract movement of the left ear, right ear, head and the body of a sow the best performing RetinaNet algorithm was applied on a data subset with 6 sows (Table 1). Originally recorded video data was downsampled from 30 fps to 1 fps to increase the speed of calculations with the RetinaNet model. Centroids of rectangles corresponding to the parts of the body of a sow were extracted and stored for each frame (Fig. 4).

In the next step Euclidean distance was calculated between centroids of the objects on each of consecutive frames,

$$d(p,q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2} \tag{1}$$

where $d(p,q)$ is the distance between points p and q and point p has Cartesian coordinates (p_1, p_2) and point q has Cartesian coordinates (q_1, q_2) . Points p and q were centroids of objects of one class (e.g. head) in consecutive frames.

In the final step the mean of euclidean distance was calculated for each object class with a sliding window of 4 h and 15 min step (Fig. 5).

2.7. Activity based on gold standard

Results of labelling of standing and walking behaviour of sows were combined into one binary variable (0, 1), which indicated if at any given second sow performed either standing or walking behaviour.

In the second step the sum of activity was calculated with a sliding window of 4 h and 15 min steps (Fig. 6).

2.8. Activity based on accelerometer data

Accelerometer data was processed and visualized as described in Oczak et al. (2019). Total physical acceleration (magnitude) was estimated from three axes of the accelerometer data (xyz) recorded in 10 Hz frequency.

The magnitude was smoothed with standard deviation calculated on a sliding window of 4 h with 15 min steps (Fig. 7).

2.9. TF models for comparison of activity measurements

Given that the expected relation between activity measurements based on ear tag accelerometer, on object detection and gold standard was assumed to be dynamic, we used the CAPTAIN Toolbox (Taylor et al., 2007) to model this relation with a Single-input, single-output (SISO) discrete transfer function model. The model has the following general structure (Young, 1984),

$$y(k) = \frac{B(Z^{-1})}{A(Z^{-1})}u(k) + \varepsilon(k) \tag{2}$$

where $y(k)$ and $u(k)$ are the output and the input of the model respectively, $\varepsilon(k)$ is additive noise assumed to be zero mean, serially uncorrelated sequence of random variables with variance σ^2 , accounting for measurement noise, modelling errors and effects of unmeasured inputs to the process; k is the sample of the measurement; $A(Z^{-1})$ and $B(Z^{-1})$ are two series given by:

$$A(Z^{-1}) = 1 + a_1 z^{-1} + a_2 z^{-2} + \ldots + a_{n_a} z^{-n_a} \tag{3}$$

$$B(Z^{-1}) = b_0 + b_1 z^{-1} + b_2 z^{-2} + \ldots + b_{n_b} z^{-n_b} \tag{4}$$

where a_j and b_j are the model parameters to be estimated; Z^{-1} is the backward shift operator, $Z^{-1}y(k) = y(k-1)$, with y and k defined as in Eq. (3) and n_a and n_b are the orders of the polynomials A and B in Eq. (4) and Eq. (5) respectively. The model parameters were estimated using a

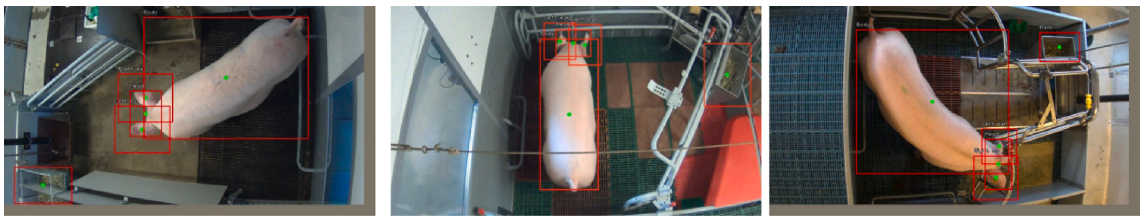


Fig. 4. Rectangles with 5 object classes in (a) SWAP pen, (b) Trapezoid pen, (c) wing pen: the right ear, left ear, body, head and the rack. Green dots indicate centroids of each object. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

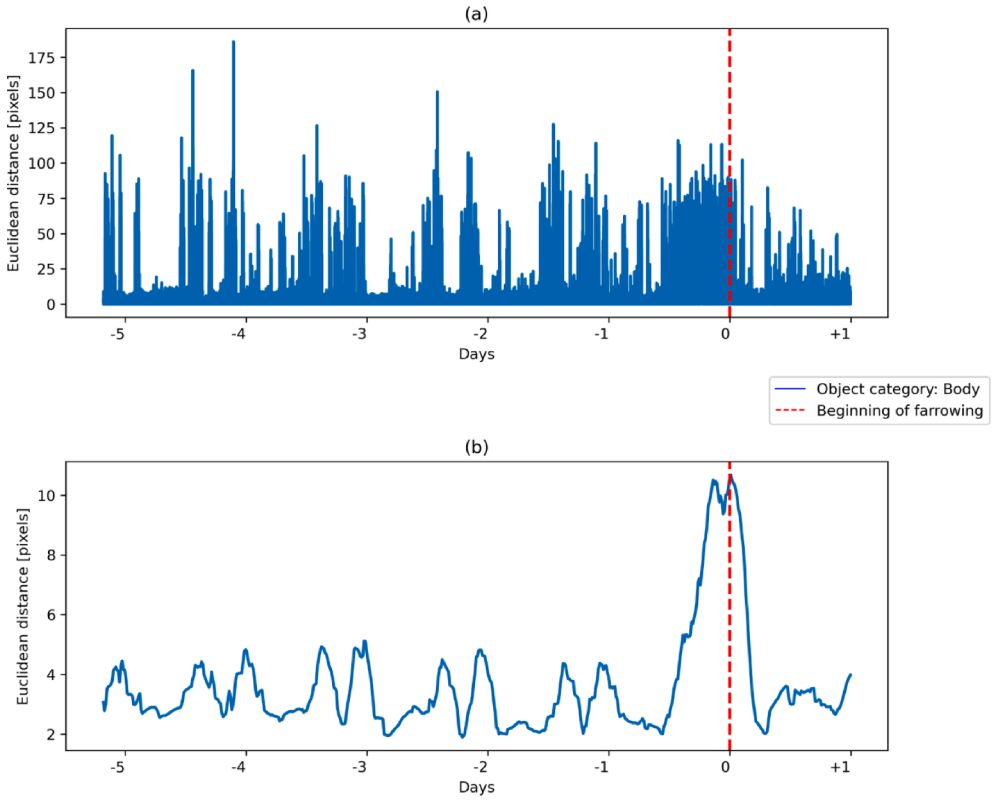


Fig. 5. Euclidean distance calculated on object class, 'body' of sow 147146–9. Period depicted in the plot starts at the introduction of a sow to a farrowing pen and ends around 1 day after beginning of farrowing. (a) Euclidean distance calculated on consecutive frames. Peak activity was a few h before the onset of farrowing. (b) Mean of Euclidean distance calculated with a sliding window of 4 h and 15 min steps. Variation in activity related to diurnal rhythms is visible.

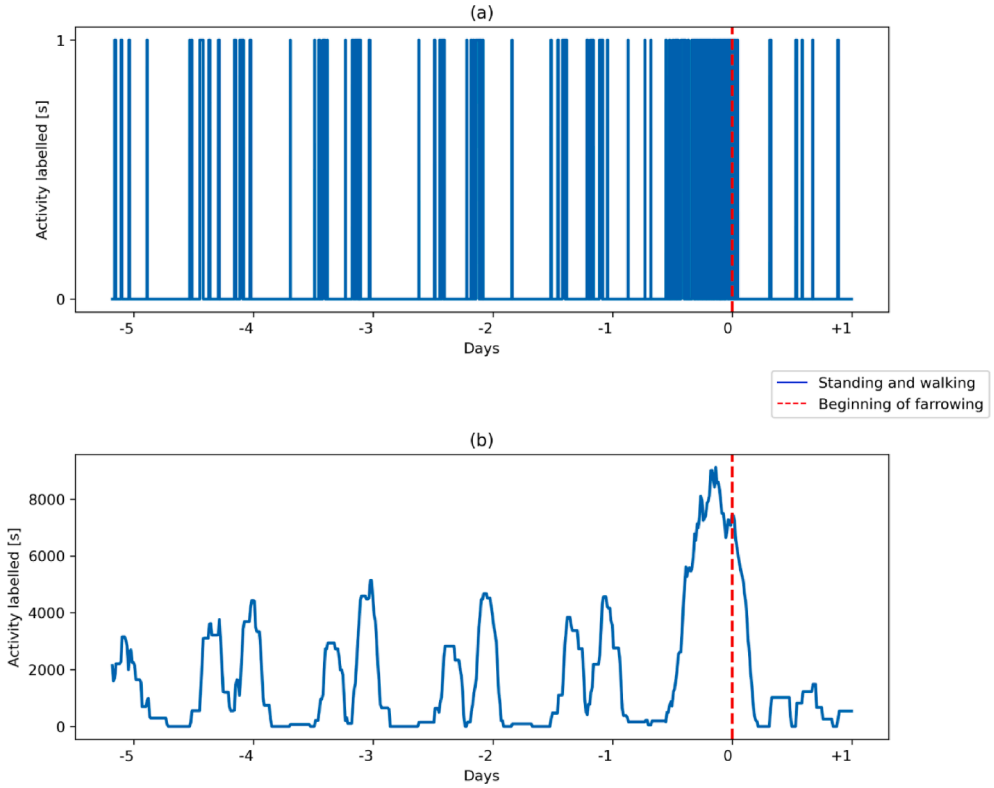


Fig. 6. Gold standard of activity of sow 147146-9. Period depicted in the plot starts at the introduction of a sow to a farrowing pen and ends around 1 day after beginning of farrowing. (a) binary variable indicating whether a sow performed standing or walking behaviour at any given second. (b) Sum calculated on labelled activity data on a sliding window of 4 h with 15 min steps. Variation in activity related to diurnal rhythm is visible.

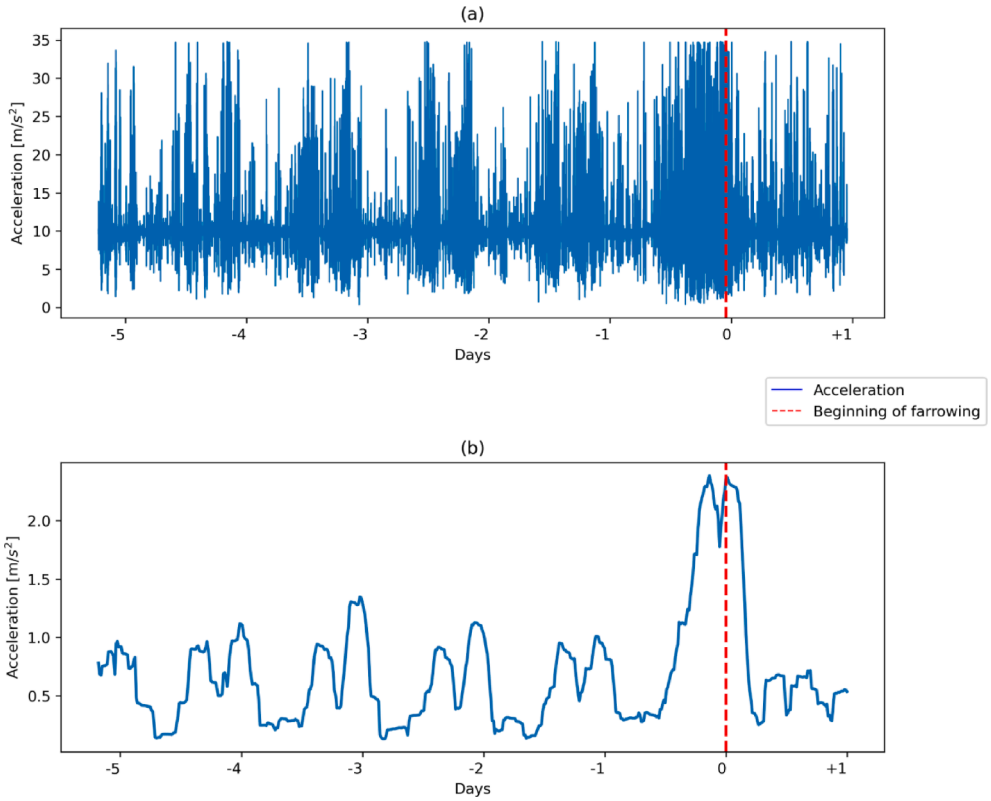


Fig. 7. Total physical acceleration of sow 147146-9. Period depicted in the plot starts at the introduction of a sow to a farrowing pen and ends around 1 day after beginning of farrowing. (a) No smoothing was applied to accelerometer data. (b) Standard deviation calculated on acceleration data on a sliding window of 4 h with 15 min steps. Variation in activity related to diurnal rhythm is visible.

refined instrumental variable approach (Young, 1984), using activity variable based on object detection as an input (either left ear, right ear, head or body) and activity variable based on ear tag accelerometer as an output. Besides, the gold standard of activity measurement was compared with activity based on computer vision by TF model with gold standard as an input variable and activity based on computer vision as an output variable. Similarly, the gold standard of activity measurement was compared with activity based on accelerometer data by TF model with gold standard as an input variable and activity based on accelerometer data as an output variable.

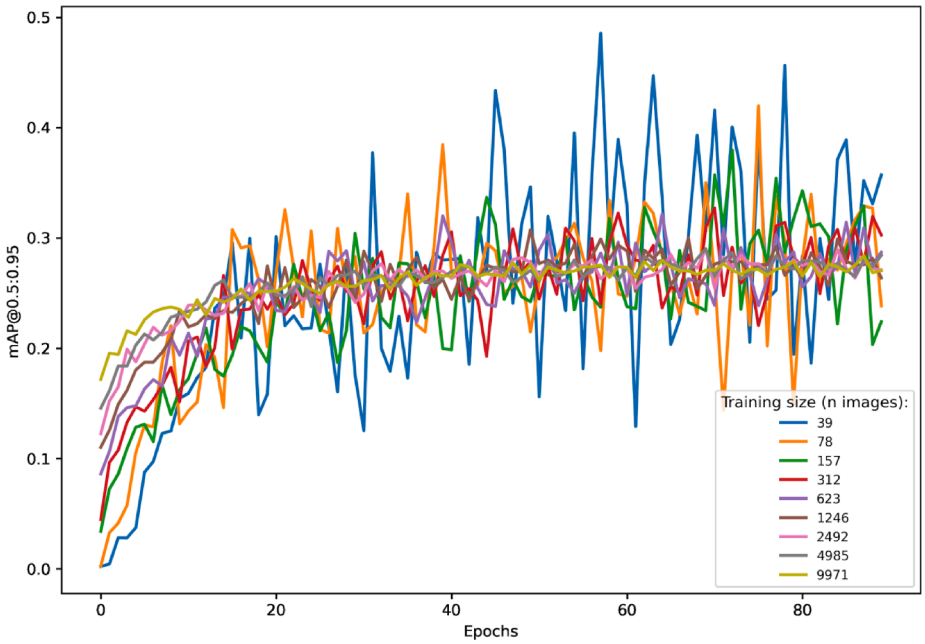


Fig. 8. Performance of RetinaNet object detector on training sets with 39, 78, 157, 312, 623, 1246, 2492, 4985 and 9971 images. Performance (mAP@0.5:0.95) is similar between different models after around 20 epochs.

3. Results

3.1. Object detection

Evaluation of performance of RetinaNet algorithms that were trained on labelled images (39–9971), indicated that mAP@0.5:0.95 on the training data sets was improving rapidly in the first 20 epochs of training (Fig. 8). Between 20 and 40 epochs there was a slight improvement of algorithm performance. Longer duration of training than 40 epochs did not improve the performance of object detection algorithms independently from the number of images used for training. In the first 20 epochs of training, algorithms trained on higher numbers of images had higher mAP@0.5:0.95 (Fig. 8).

Validation of RetinaNet algorithms on 4273 images indicated that mAP@0.5:0.95 was consistently higher for the algorithms trained on a larger number of images. Improvement in the performance of object detectors, as on the training set, is clear in the first 20 epochs and slight between 20 and 40 epochs. Longer training than 40 epochs did not improve the performance of the algorithms on the validation set (Fig. 9a). Performance of RetinaNet models on the validation set increased by around 0.02 mAP@0.5:0.95 as the size of the training set was doubled (Fig. 9b).

These results suggest that further increase of the number of frames in the training set might further improve the performance of the object detector on the validation set. Although, effort related to labelling of more frames to achieve improvement of 0.02 mAP@0.5:0.95 would grow 2 fold. Thus, increasing the performance by 0.02 mAP@0.5:0.95 above the maximum performance that we achieved with 9971 images in the training set would probably require labelling twice the number of images (19942) (see Fig. 10).

The best performance of RetinaNet was achieved for the biggest object in the pen, the body of a sow, 0.37 AP@0.5:0.95. Hay rack was detected with AP@0.5:0.95 of 0.26 which indicated worse performance of the object detector than for the body of a sow.

Hay rack was also smaller than the body of a sow and of similar size to the head of a sow. Rack was the only object that was not moving in the pen. Slightly worse performance in comparison to a rack was achieved for a head (0.25 AP@0.5:0.95). The worst performance was achieved for the smallest objects that we focused on, the left and right ears of a sow (0.22 AP@0.5:0.95) (Table 2). Performance of RetinaNet on labelled objects was consistently improving as the size of the training set was increased with the best performance achieved for the body of a sow and

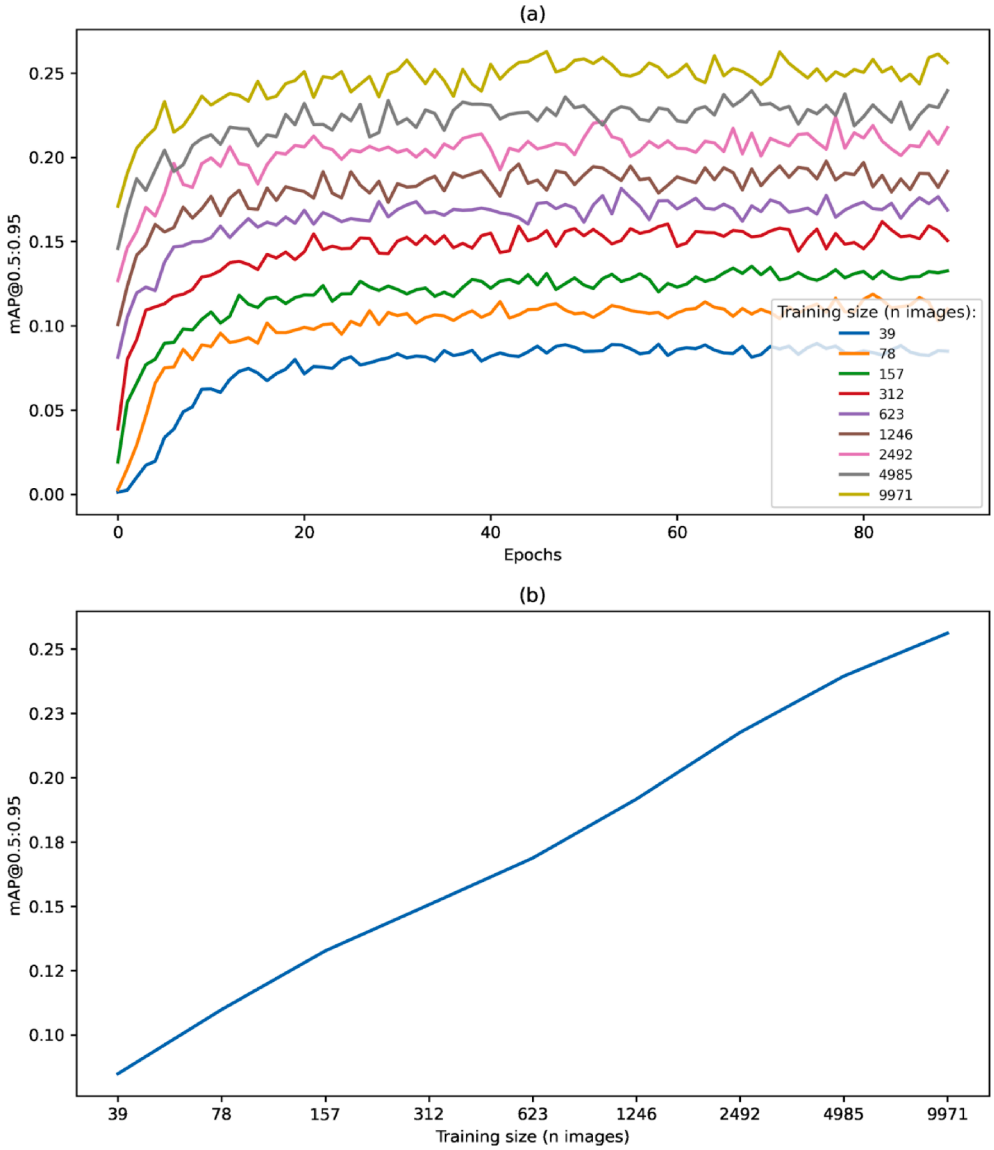


Fig. 9. Performance of RetinaNet object detector on a validation set with 4273 images. (a) Models trained on more images have higher mAP@0.5:0.95 (b) Linear improvement in mAP@0.5:0.95 of models trained for 90 epochs.

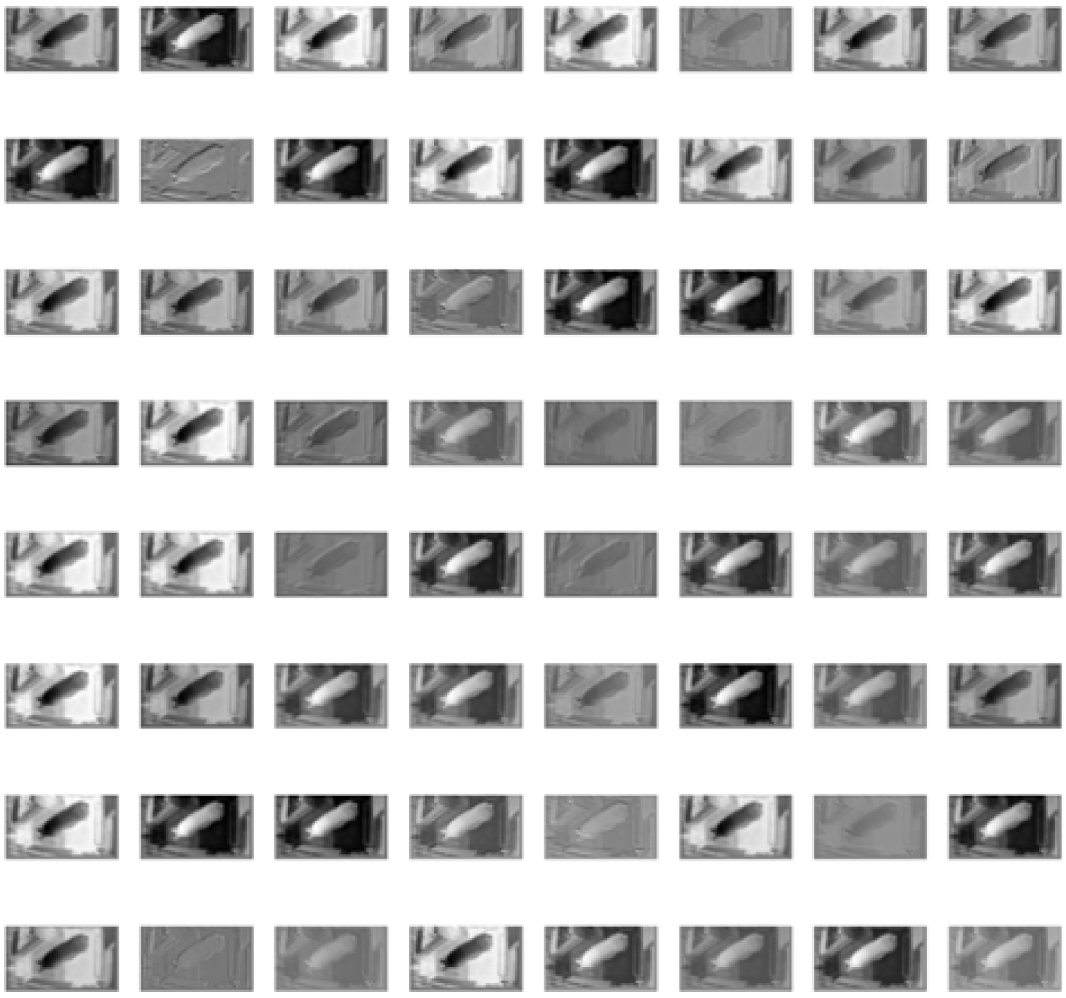


Fig. 10. Feature map with 64 features from 33rd layer of RetinaNet model.

Table 2
Performance of object detection models validated on 4273 images.

Object class	Training size (n images)								
	39	78	157	312	623	1246	2492	4985	9971
Left ear (AP@0.5:0.95)	0.02	0.03	0.05	0.07	0.09	0.12	0.16	0.19	0.22
Right ear (AP@0.5:0.95)	0.01	0.03	0.05	0.07	0.08	0.11	0.16	0.19	0.22
Body (AP@0.5:0.95)	0.22	0.26	0.26	0.32	0.32	0.33	0.33	0.34	0.37
Hay rack (AP@0.5:0.95)	0.12	0.15	0.17	0.22	0.22	0.23	0.24	0.24	0.26
Head (AP@0.5:0.95)	0.05	0.08	0.11	0.14	0.16	0.17	0.19	0.23	0.25
All (mAP@0.5:0.95)	0.08	0.11	0.13	0.15	0.17	0.19	0.22	0.24	0.26

the worst for the ears (see Table 3).

The most important results of application of the RetinaNet algorithm was that training the algorithm for longer than 40 epochs didn't improve its performance. Additionally, bigger objects such as sow bodies are detected with better performance than smaller objects such as ears or heads of sows. To our surprise the hay rack, which was not moving in the pen, was detected with worse performance than the sow's body which was changing location in the pen and with similar performance to the head of the sow. This suggests that the size of the object might be more important for the performance of object detection algorithms than other factors such as movement of these objects between frames. Performance of RetinaNet trained on labelled images (39–9971) is presented in supplementary video material.

Table 3
Additional metrics of object detection models trained on 9971 and validated on 4273 images.

Object class	AP@0.5:0.95	AP@0.5	AP@0.75
Left ear	0.22	0.28	0.23
Right ear	0.22	0.28	0.24
Body	0.37	0.49	0.36
Hay rack	0.26	0.27	0.26
Head	0.25	0.31	0.28

3.2. Activity

Fitting agreement (R^2) of TF models between gold standard based on human labelling of active behaviour and object detector RetinaNet, indicated that for 5 out of 6 animals gold standard was in high agreement with activity based on detection of the whole body of sows

($R^2 > 0.8$) (Table 4). Mean R^2 between gold standard and activity based on detection of the whole body of sows was the highest ($R^2 = 0.84$). It was the lowest between gold standard and activity based on detection of the left ear ($R^2 = 0.53$). Mean R^2 between gold standard and activity based on detection of heads of sows was higher than for ears and lower than for the whole body of sows ($R^2 = 0.75$) (Table 4).

Mean R^2 between gold standard and activity based on detection of the whole body of sows was higher in SWAP ($R^2 = 0.89$) and Wing pens ($R^2 = 0.87$) than trapezoid pens ($R^2 = 0.75$) (Table 4).

R^2 of TF models between gold standard and ear tag accelerometer data, also indicated that for the same 5 out of 6 animals gold standard was in high agreement with activity based on ear tag accelerometer data ($R^2 > 0.8$) (Table 5). Mean R^2 between gold standard and ear tag accelerometer data was higher in Wing pens ($R^2 = 0.87$) than in SWAP ($R^2 = 0.80$) and trapezoid pens ($R^2 = 0.78$) (Table 5).

Comparison of R^2 of TF models between activity measurements based on object detection and accelerometer data revealed a similar pattern as between activity based on gold standard, object detector and accelerometer data. Mean R^2 was the highest between accelerometer based activity and the whole body of sows ($R^2 = 0.81$) and the lowest between accelerometer based activity and left ears ($R^2 = 0.46$) (Table 6).

R^2 between activity measurements based on object detection for both ears and accelerometer data was mostly lower than for the head and the body (Table 6). Except for two animals R^2 was below 0.65 for the left and/or the right ear. For one of the animals (147002 7) fitting agreement of TF models between gold standard and object detection was high for all body parts ($R^2 > 0.8$). Activity measurements for this animal are presented in Fig. 11. Mean R^2 between activity measurements based on object detection for the whole body of sows and ear tag accelerometer data was higher in Wing pens ($R^2 = 0.91$) than in SWAP ($R^2 = 0.75$) and trapezoid pens ($R^2 = 0.76$) (Table 6).

For animal 147038 9 R^2 of TF models between gold standard and object detection of sows body was lower ($R^2 = 0.58$) for the other sows and this indicated poor performance of object detection on this animal (Fig. 12).

High R^2 between activity measurement based on accelerometer data and based on head and body and gold standard suggests that if performance of object detection is good enough to provide a continuous reading the measurements of activity are very similar between both sensors, a camera and an accelerometer.

Calculation of animals' activity based on automated detection of location of sows' ears was not as reliable as based on location of heads and the whole body. This is probably due to the fact that ears are smaller than heads or the whole body of a sow and were more prone to occlusions by other parts of the sow body or pen elements such as a crate. These results suggest that detection of the whole body of a sow or heads of the animal is a good basis for estimating the animal's activity with computer vision but other parts of the sow's body such as ears cannot provide a reliable estimate of animal's activity.

4. Discussion

Object detection based on Deep Neural Networks (DNNs) has been applied so far to automatically classify postures of sows in a farrowing compartment (Zheng et al., 2018), to detect fattening pigs and their heads in a group pen (Yang et al., 2018), to detect both body part locations and pairwise associations of the body parts of fattening pigs (Psota et al., 2019), to detect fattening pigs in a group pen (Liu et al., 2020) and to detect enrichment objects in a fattening pen (Chen et al., 2020). However, to our knowledge it has been the first time that object detection algorithm has been applied to detect parts of the body of a sow housed in a farrowing pen. Presence of a farrowing crate in a pen caused an additional challenge for object detection algorithms as parts of the body of a sow were often occluded by the farrowing crate. Bounding box representation used in object detection algorithms implicitly assumes that all pixels inside the box belong to the object. This assumption makes this representation less robust to the object with occlusion (Dollar et al., 2009).

Our results of object detection are impossible to compare with all of the results of object detection reported in above referenced articles as no metric of object detection performance is mostly reported. Moreover, the lack of consensus in different works and AP implementations is a problem faced by the academic and scientific communities (Padilla et al., 2020). In research of Liu et al. (2020) mAP@0.75 of 1 is reported for detection of fattening pigs in a group pen, which could be compared with 0.36 AP@0.75 for detection of a body of a sow in our research. Much better performance of detection of fattening pigs in a group pen might be related to the fact that only 64 images were used for algorithm validation in the research of Liu et al. (2020) and all images were recorded in the same fattening pen. In our research 4273 images were used as a validation set and were recorded in 3 types of farrowing pens. This higher variability of pen structures together with frequent occlusions of the animals possibly reduced the overall performance of object detection.

In research of Liu et al. (2020), Zheng et al. (2018) and Chen et al. (2020) a transfer learning approach was used. Transfer learning allows use of models pre-trained on big datasets for other tasks without a necessity to train a model from scratch, but rather to fine-tune an existing model with own datasets (Pan and Yang, 2010). According to Liu et al. (2020) transfer learning greatly improved the efficiency of training an object detection model in their research as otherwise it might take weeks and millions of images to fit a deep model if training from scratch. In our research we used a pretrained ResNet backbone and fine-tuned it with

Table 4
R^2 between the gold standard for activity measurement and on object detection.

Animal id	Pen type	R2			
		Body	Head	Left ear	Right ear
147002 7	SWAP	0.93	0.90	0.83	0.80
147127 10	SWAP	0.85	0.87	0.82	0.67
147038 9	Trapezoid	0.58	0.65	0.06	0.16
147142 2	Trapezoid	0.93	0.92	0.58	0.88
147142 10	Wing	0.91	0.91	0.83	0.78
147146 9	Wing	0.84	0.24	0.09	0.22
Mean	–	0.84	0.75	0.53	0.58

Table 5
R^2 between the gold standard for activity measurement and activity based on ear tag accelerometer data.

Animal id	Pen type	R2 Acc
147002 7	SWAP	0.80
147127 10	SWAP	0.80
147038 9	Trapezoid	0.64
147142 2	Trapezoid	0.93
147142 10	Wing	0.82
147146 9	Wing	0.92
Mean	–	0.82

Table 6
R^2 between activity based on ear tag accelerometer data and on object detection.

Animal id	Pen type	R2			
		Body	Head	Left ear	Right ear
147002 7	SWAP	0.79	0.76	0.50	0.60
147127 10	SWAP	0.71	0.75	0.63	0.29
147038 9	Trapezoid	0.70	0.60	0.19	0.41
147142 2	Trapezoid	0.82	0.80	0.48	0.81
147142 10	Wing	0.95	0.83	0.76	0.84
147146 9	Wing	0.88	0.37	0.20	0.22
Mean	–	0.81	0.69	0.46	0.53

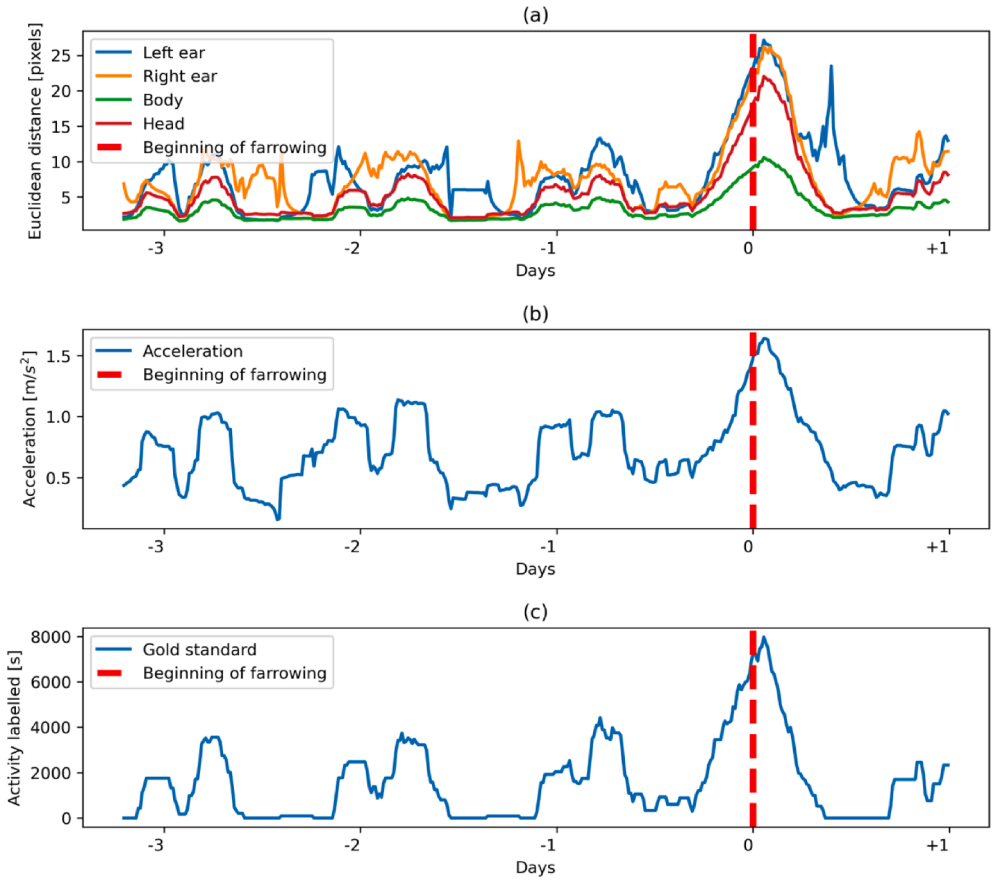


Fig. 11. Activity of a sow 147002 7 measured based on object detection: (a) Left ear, right ear, body, head. (b) Activity based on ear tag accelerometer. (c) Gold standard based on human labelling.

9969 images within around 3 days on a single Titan RTX GPU. It was possible to train our model in such a short period of time because we used the transfer learning approach.

Both in the research of Liu et al. (2020) and Chen et al. (2020) output of object detection algorithms was used as a basis for tracking algorithms applied in the second stage of extracting information from video data. Application of k-means algorithm in our research to select frames for training and validation of object detection, on the one hand allowed more robust training and validation of RetinaNet with minimal number of images. On the other hand images in the validation set were not sequences of consecutive video frames and this prevents any improvements in performance of object detection when tracking algorithms would have been applied to those images. However, application of tracking algorithm might have improved the performance of object detection when applied to consecutive video frames in the data subset for comparison of computer vision approach with accelerometer. In future research we plan to compare the performance of object detection algorithms with object detection algorithms and added tracking (e.g. Simple Online and Realtime Tracking with a Deep Association Metric). In such research the validation dataset could be composed of two subsets, one with frames selected by k-means algorithm and a second one with labelled consecutive sequences of images. Application of a tracking algorithm might improve the general performance of object detection and it might also reduce the number of images that need to be labelled to train the algorithm.

Comparison of measurement of activity of sows in the farrowing compartment between computer vision and ear tag accelerometer revealed that activity measurement based on object classes on which RetinaNet had better performance such as head or body of a sow had higher R^2 with activity measurement based on accelerometer data and gold standard based on human labelling. It seems that measurement of activity of an animal based on a computer vision approach represents activity of the whole body of the sow or head very well as long as object detection also had good performance. We compared measurement of activity of a sow with both technologies in our research mainly to explore a potential application of computer vision for farrowing prediction as an extension of using accelerometer technology for this purpose (Oczak et al., 2019). Thus, we applied a smoothing window of 4 h with 15 min steps to activity measurement based on object detection and this allowed us to monitor a diurnal rhythm of a sow and also shifts of activity level as the farrowing approached. High fitting agreement of activity measurement based on both technologies suggests that there might be no direct added value of using measurement of activity level based on computer vision for farrowing prediction, except for non invasive character of computer vision approach. It is possible that differences between measurements based on both technologies would be more apparent if raw data or a shorter smoothing window than 4 h is used.

In the future research on farrowing prediction we will rather focus on automated detection of specific nest-building behaviours such as usage of a hay rack in a farrowing pen by application of computer vision methods. Such an approach might have advantages in comparison to application of accelerometer data as proximity to different objects in the pen can be estimated with computer vision techniques whereas based on accelerometer data this cannot be done.

Considering possible application of both presented methods for monitoring pig activity in different environments (e.g multiple farms) we included 3 types of farrowing pens in our experimental setup. Comparison of R^2 of both presented methods with gold standard between 3 pen types revealed that R^2 varied between pens for both

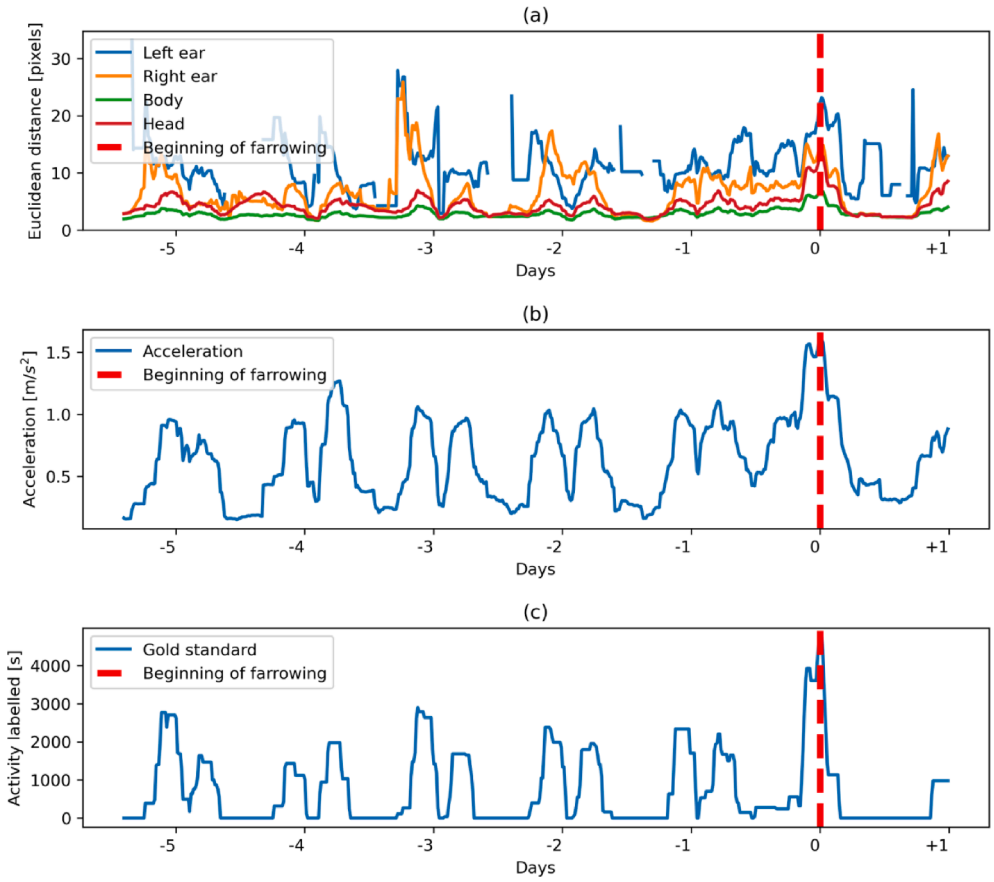


Fig. 12. Activity of a sow 147038 9 measured based on object detection: (a) Left ear, right ear, body, head. Missing data is visible for all variables based on object detection. (b) Activity based on ear tag accelerometer. (c) Gold standard based on human labelling.

methods with lowest R^2 for Trapezoid pens for both methods and highest for SWAP (computer vision) and Wing pens (ear tag accelerometer). This result suggests that the environment might have an impact on performance of presented methods but with low sample size drawing clear conclusions which method might be more robust when applied in new environments is not possible. What has to be considered in this context is that the performance of a method for monitoring pig activity based on video data presented in this publication depends on performance of RetinaNet object detection algorithm. RetinaNet has been trained and validated on the same three pen types in our work SWAP, trapezoid and wing. Performance of another deep learning object detection algorithm YOLO (You Only Look Once) trained and validated for detection of fattening pigs dropped significantly when trained on one farm and validated on another and mAP didn't even reach 0.01 (Van Caenegem, 2020). This was due to various lighting conditions or positions of the cameras between farms. In research of Van Caenegem (2020) 40 to 60 images collected on a second farm were needed to achieve decent performance of YOLO in the new environment. We expect that re-training of the RetinaNet object detection algorithm with a similar number of images on additional farms would be needed to achieve comparable performance of our method for activity monitoring in new environments. We expect that method for activity monitoring based on ear tag accelerometer data presented in our publication is more robust when validated on additional farms as the same movement of the sensor on another farm will provide the same estimate of animal activity. This might not be the case when the RetinaNet object detection algorithm is not re-trained with additional images from other farms due to low mAP of the algorithm on additional farms meaning inability of the model to correctly detect sow body parts or false detections.

Our comparison of measurement of activity of sows in the farrowing compartment between computer vision and ear tag accelerometer revealed that non-invasive method based on video data can provide very similar information on activity level of animals to invasive (ear pinching) and laborious (regular battery exchange) ear tag accelerometer based method. The presented computer vision method is limited to monitoring one animal under camera view as detected body parts cannot be associated with multiple individuals. Moreover, we expect that the method requires re-training the RetinaNet object detection algorithm with additional images collected on additional farms to achieve satisfactory performance in different environments.

5. Conclusions

In this research RetinaNet object detection algorithm was applied to automatically detect parts of the body of a sow and a hay rack in a farrowing pen. It was possible to detect these objects with a performance of 0.26 mAP@0.5:0.95. On the basis of output of RetinaNet activity of sows was calculated, smoothed over a window of 4 h and compared with activity measurement based on ear tag accelerometer data. Results of fitting of TF models to animal activity data based on ear tag accelerometer and output of object detection on body of sows and head of sows suggests that both technologies, the accelerometer and camera provide very similar information on activity level of animals, considering that object detection algorithm was trained and validated on the same farm.

CRediT authorship contribution statement

Maciej Oczak: Conceptualization, Methodology, Software, Writing – original draft. **Florian Bayer:** Writing – review & editing. **Sebastian Vetter:** Writing – review & editing. **Kristina Maschat:** Resources, Data

curation. **Johannes Baumgartner:** Supervision, Funding acquisition.

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.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compag.2021.106517.

References

Ala-Kurikka, E., Heinonen, M., Mustonen, K., Peltoniemi, O., Raekallio, M., Vainio, O., Valros, A., 2017. Behavior changes associated with lameness in sows. Appl. Anim. Behav. Sci. 193, 15–20.

Baxter, E.M., Jarvis, S., Sherwood, L., Farish, M., Roehe, R., Lawrence, A.B., Edwards, S. A., 2011. Genetic and environmental effects on piglet survival and maternal behaviour of the farrowing sow. Appl. Anim. Behav. Sci. 130 (1-2), 28–41.

Castrén, H., Algers, B., de Passillé, A.-M., Rushen, J., Uvnäs-Moberg, K., 1993. Preparturient variation in progesterone, prolactin, oxytocin and somatostatin in relation to nest building in sows. Appl. Anim. Behav. Sci. 38 (2), 91–102.

Chapa, J.M., Maschat, K., Iwersen, M., Baumgartner, J., Drillich, M., 2020. Accelerometer systems as tools for health and welfare assessment in cattle and pigs – a review. Behav. Process. 181, 104262. https://doi.org/10.1016/j.beproc.2020.104262.

Chen, C., Zhu, W., Oczak, M., Maschat, K., Baumgartner, J., Larsen, M.L.V., Norton, T., 2020. A computer vision approach for recognition of the engagement of pigs with different enrichment objects. Comput. Electron. Agric. 175, 105580. https://doi.org/10.1016/j.compag.2020.105580.

Contreras-Aguilar, M.D., Escribano, D., Martínez-Miró, S., López-Arjona, M., Rubio, C.P., Martínez-Subiela, S., Cerón, J.J., Tecles, F., 2019. Application of a score for evaluation of pain, distress and discomfort in pigs with lameness and prolapses: correlation with saliva biomarkers and severity of the disease. Res. Vet. Sci. 126, 155–163.

Cornou, C., Lundbye-Christensen, S., 2008. Classifying sows' activity types from acceleration patterns. Appl. Animal Behav. Sci. 111 (3-4), 262–273.

Costa, A., Borgonovo, F., Leroy, T., Berckmans, D., Guarino, M., 2009. Dust concentration variation in relation to animal activity in a pig barn. Biosystems Eng. 104 (1), 118–124.

Dollar, P., Wojek, C., Schiele, B., Perona, P., 2009. Pedestrian detection: A benchmark, in: 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp. 304–311.

Erez, B., Hartsock, T.G., 1990. A microcomputer-photocell system to monitor periparturient activity of sows and transfer data to remote location. J. Anim. Sci. 68, 88–94.

Hamäläinen, W., Järvinen, M., Martiskainen, P., Mononen, J., 2011. Jerk-based feature extraction for robust activity recognition from acceleration data. In: 2011 11th International Conference on Intelligent Systems Design and Applications, pp. 831–836.

Hart, B.L., 1988. Biological basis of the behavior of sick animals. Neurosci. Biobehav. Rev. 12 (2), 123–137.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778.

Heinonen, M., Peltoniemi, O., Valros, A., 2013. Impact of lameness and claw lesions in sows on welfare, health and production. Livestock Sci. 156 (1–3), 2–9.

Huynh, T.T.T., Aarnink, A.J.A., Verstegen, M.W.A., 2005. Reactions of pigs to a hot environment. Presented at the Livestock Environment VII, American Society of Agricultural and Biological Engineers, pp. 544.

Jensen, P., 1986. Observations on the maternal behaviour of free-ranging domestic pigs. Appl. Anim. Behav. Sci. 16 (2), 131–142.

Johnson, A.K., Morrow, J.L., Dailey, J.W., McGlone, J.J., 2007. Preweaning mortality in loose-housed lactating sows: Behavioral and performance differences between sows who crush or do not crush piglets. Appl. Anim. Behav. Sci. 105 (1-3), 59–74.

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.

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L., 2014. Microsoft COCO: Common Objects in Context, in: Computer Vision – ECCV 2014. Springer International Publishing, pp. 740–755.

Liu, D., Oczak, M., Maschat, K., Baumgartner, J., Pletzer, B., He, D., Norton, T., 2020. A computer vision-based method for spatial-temporal action recognition of tail-biting behaviour in group-housed pigs. Biosyst. Eng. 195, 27–41.

Lopes, P.C., 2017. Why are behavioral and immune traits linked? Horm. Behav. 88, 52–59.

Martínez-Avilés, M., Fernández-Carrión, E., López García-Baones, J.M., Sánchez-Vizcaíno, J.M., 2017. Early detection of infection in pigs through an online monitoring system. Transbound. Emerg. Dis. 64 (2), 364–373.

Oczak, M., Maschat, K., Baumgartner, J., 2019. Dynamics of sows' activity housed in farrowing pens with possibility of temporary crating might indicate the time when sows should be confined in a crate before the onset of farrowing. Animals 10 (1), 6. https://doi.org/10.3390/ani10010006.

Oliviero, C., Pastell, M., Heinonen, M., Heikkonen, J., Valros, A., Ahokas, J., Vainio, O., Peltoniemi, O.A.T., 2008. Using movement sensors to detect the onset of farrowing. Biosyst. Eng. 100 (2), 281–285.

Padilla, R., Netto, S.L., da Silva, E.A.B., 2020. A survey on performance metrics for object-detection algorithms. In: 2020 International Conference on Systems, Signals and Image Processing (IWSSIP), pp. 237–242.

Pan, S.J., Yang, Q., 2010. A survey on transfer learning. IEEE Trans. Knowl. Data Eng. 22 (10), 1345–1359.

Pereira, T.D., Aldarondo, D.E., Willmore, L., Kislin, M., Wang, S.-H., Murthy, M., Shaevitz, J.W., 2019. Fast animal pose estimation using deep neural networks. Nat. Methods 16 (1), 117–125.

Psota, E., Mittek, M., Pérez, L., Schmidt, T.y., Mote, B., 2019. Multi-pig part detection and association with a fully-convolutional network. Sensors 19 (4), 852. https://doi.org/10.3390/s19040852.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A.C., Fei-Fei, L.i., 2015. ImageNet large scale visual recognition challenge. Int. J. Comput. Vis. 115 (3), 211–252.

Taylor, C., Pedregal, D., Young, P., Tych, W., 2007. Environmental time series analysis and forecasting with the Captain toolbox. Environ. Modell. Softw. 22 (6), 797–814.

Van Caenegem, R., 2020. Detecting and Locating Individual Pigs from the Group Using Deep Learning (MSc). KU Leuven, Leuven.

Von Jasmund, N., Wellnitz, A., Krommweh, M.S., Büscher, W., 2020. Using passive infrared detectors to record group activity and activity in certain focus areas in fattening pigs. Animals (Basel) 10 (5), 792. https://doi.org/10.3390/ani10050792.

Wathes, C.M., Kristensen, H.H., Aerts, J.-M., Berckmans, D., 2008. Is precision livestock farming an engineer's daydream or nightmare, an animal's friend or foe, and a farmer's panacea or pitfall? Comput. Electron. Agric. 64 (1), 2–10.

Weary, D.M., Huzzey, J.M., von Keyserlingk, M.A.G., 2009. Board-invited review: using behavior to predict and identify ill health in animals. J. Anim. Sci. 87, 770–777.

Yang, Q., Xiao, D., Lin, S., 2018. Feeding behavior recognition for group-housed pigs with the Faster R-CNN. Comput. Electron. Agric. 155, 453–460.

Young, P. (Ed.), 1984. Recursive Estimation and Time-Series Analysis. Springer Berlin Heidelberg, Berlin, Heidelberg.

Zheng, C., Zhu, X., Yang, X., Wang, L., Tu, S., Xue, Y., 2018. Automatic recognition of lactating sow postures from depth images by deep learning detector. Comput. Electron. Agric. 147, 51–63.