# A PCA-based frame selection method for applying CNN and LSTM to classify postural behaviour in sows

Meiqing Wang [a], Maciek oczak [b,c], Mona Larsen [a], Florian Bayer [b], Kristina Maschat [b], Johannes Baumgartner [b], Jean-Loup Rault [b], Tomas Norton [a,*]

[a] Faculty of Bioscience Engineering, Katholieke Universiteit Leuven (KU LEUVEN), Kasteelpark Arenberg 30, 3001 Heverlee/Leuven, Belgium
[b] Institute of Animal Welfare Science (ITT), University of Veterinary Medicine (Vetmeduni) Vienna, Veterinaerplatz1, A-1210 Vienna, Austria
[c] Precision Livestock Farming Hub, The University of Veterinary Medicine Vienna (Vetmeduni Vienna), Veterinärplatz 1, 1210 Vienna, Austria

ARTICLE INFO

*Keywords:*
Behaviour classification
PCA
Frame selection
Video analysis
Pig
Welfare

ABSTRACT

Posture and the rate of postural changes of farrowing and lactating sows are considered reliable indicators of environmental comfort and health status and are risk factors for piglet crushing. The objective of this study was to develop a combined deep learning and principle component analysis (PCA) based approach to classify different postural behaviours of sows in videos. Compared to previous studies of sow's postural behaviour classification based on deep learning, this study selects sequences of frames from the videos that distinguish different postural behaviours rather than using all frames for the classification. Videos were collected from 13 sows, and the recording started from 5 days before the expected date of farrowing until weaning. From the videos, 3100 videos without piglets and 1680 including piglets were manually selected. Then, these videos were augmented by using vertical mirroring and adding Gaussian noise, which resulted in 7200 and 4600 videos without and including piglets, respectively. Each video lasted 5 sec and included 1 out of 5 behavioural postures (sternal lying, lateral lying, sitting, standing, walking) labelled by one trained expert with extensive experience in sow's behaviour classification. Out of the total of 11,800 videos, 75% were randomly allocated as training set and the remaining 25% as validation set. To select motion-related frames, each video was first converted into a multidimensional matrix. Then, PCA was performed on the matrix and a number of component(s) were selected to represent the frame. After that, the frame Euclidean distances were computed based on the components and the frames over a certain distance threshold were selected to generate new videos. Since a different number of components and distance thresholds can affect the number of selected frames, a range of component numbers (1, 2, 3, 5, 10, 20, 50) and distance thresholds were further tested to find the optimal parameters. The best balance between accuracy and performance of the classification was obtained when using 10 components (87.98% of total variation). The best results were obtained when the threshold was set as one fourth of the largest distance between two successive frames. To classify different behaviours, the videos composed of the selected frames were trained and validated with convolutional neural network (CNN) and a long short-term memory (LSTM) models. Using the proposed method, postural behaviours could be classified with accuracies of 95.33% and 92.67% on videos without piglets and all data (including and not including piglets). Furthermore, 500 new videos were selected from the experiment and were used as test set. The final model was further tested on the test set and returned an accuracy of 90.60%, which indicated that the proposed method can be generalized on new data.

## 1. Introduction

The postural behaviour of sows can be an indicator of their welfare and health. For instance, heat stress induced by ambient temperature changes cause different postural changes in sows. Sows lie down laterally with their limbs extended when the ambient temperature is high, whereas they lie down sternally, which minimizes their contact with the floor, in low environmental temperatures (Huynh et al., 2005; Spoolder

et al., 2012). Lameness, as a prevalent health and welfare concern leading to shorter longevity and economic loss, can also be assessed indirectly by quantifying postural behaviours. Lame sows spend less time standing and lay down earlier after feeding than non-lame sows (Grégoire et al., 2013). Lame sows are also more likely to show difficulties in lying down than healthy individuals (Bonde et al., 2004). Shoulder sores can be caused by prolonged lying on inappropriate floor due to inadequate feeding and can also lead to behavioural changes. Larsen et al. (2015) showed that sows with shoulder sores spent less time lying, tended to perform more postural changes, spent more time standing still, and showed increased shoulder rubbing and reduced nursing frequency compared to healthy individuals (Larsen et al., 2015).

Apart from the welfare and health of the sow itself, the postural behaviour of sows is also related to productivity in pig farming. For instance, the postural transition from "sitting to lying" and the movement change from "standing to lying" as well as rolling behaviour (changing lateral lying position from one side to the other) are important causes of piglet crushing in the first 72 h after birth (Nicolaisen et al., 2019; Damm et al., 2005), with piglet crushing as the main cause of piglet mortality in indoor housing systems (Marchant et al., 2001). Sows become more active before farrowing due to nest-building behaviour and the analysis of behaviour patterns may indicate the timing of parturition and the need for closer monitoring by the farmer (Oczak et al., 2020). Compared to traditional human observation of animals, computer vision methods are more time-efficient given that farmers can get the postural information directly and automatically from real-time video analysis. Computer vision methods are also non-intrusive and non-stressful for animals compared to accelerometers or other body-mounted/wearable sensors. Therefore, there has been great interest in using computer vision methods for classifying different animal behaviours.

Application of computer vision to automatically classify the behaviour of pigs was initiated by Kashihaand colleagues who used ellipse fitting to localize the pig and then computed the amount of moving pixels in each image to determine pigs as active or inactive (Kashiha et al., 2013). Nasirahmadi *et al.* scored the lateral and sternal lying postures by calculating the area and perimeter of the boundary and convex hull on single images, and classified them using a support vector machine (SVM) (Nasirahmadi et al., 2019). However, the classification accuracies found in the above studies can easily be affected by the image quality and the accuracy of pixels or points detected on the sow's body. Deep learning techniques can help to deal with the problem of low image quality and pixel/point detection. For instance, a fully convolutional network (FCN) was developed for lactating sow image segmentation with different image qualities (Yang et al., 2018). Moreover, Zheng *et al.* developed a Faster region-based convolutional neural networks (Faster R-CNN) model to classify five postural behaviours (standing, sitting, sternal recumbency, ventral recumbency and lateral recumbency) in sows (Zheng et al., 2018).

However, the above studies only extracted spatial features from still images to classify the sow's behaviours, which cannot simultaneously obtain the coherent temporal information of the behaviours. The temporal information between consecutive frames is important, especially when distinguishing between active and inactive behaviours, *e.g.* walking *vs.* standing. Furthermore, the rapid development of deep learning techniques brought new opportunities to behavioural classification in videos. For instance, a two-stream convolutional network model was developed to extract the temporal and spatial features based on video analysis to classify five behaviours (feeding, lying, walking, scratching, mounting) in pigs (Zhang et al., 2020). Li *et al.* proposed a spatiotemporal convolutional network that can extract different features from low and high frame rate videos to classify different behaviours of pigs (Li et al., 2020). Most published methods analysed an entire video and used all frames to do the classification. However, human vision proves that simple actions can be recognized almost instantaneously (Schindler and Van Gool, 2008). The behaviour of sows could therefore potentially be correctly recognized from very short sequences, which indicates that extracting the features from all frames may be using more information than required. Additionally, processing the entire video is time-consuming compared to only processing motion-related frames, *i.e.* the frames that show the main movements of the animal and can most greatly distinguish different postural behaviours. Specifically, spatial and temporal features need to be extracted from all frames when using the entire video, whereas the processing of motion-related frames only needs to deal with a small part of the video while potentially being able to retain good performance.

The current study attempted to select motion-related frames from the video and then classify them into different behaviours based on the selected frames. Principle component analysis (PCA) is a reliable tool for dimension reduction and has already been used in many cases of animal image analysis, *e.g.* pig detection (Sun et al., 2019) and fish detection and recognition (Matai et al., 2012). This study aimed to adopt PCA to reduce the dimensions of frame and to represent the frame by chooseing a number of component(s). Compared to other cluster-based frame selection methods (Zhuang et al., 1998; Ferman and Tekalp, 1997) identifying cluster centers as key frames, PCA-based method don't need to iteratively compute the cluster centroids and are more applicable to less complex scenarios. Additionally, CNN along with LSTM were shown to be reliable for extracting spatial–temporal information for pig aggressive behaviour detection (Chen et al., 2020), tail-biting behaviour recognition (Liu et al., 2020), as well as drinking and drinker-playing behaviour classification (Chen et al., 2020). Therefore, the objective of this study was to develop a PCA based frame selection method for applying CNN and LSTM to classify sows' postural behaviour.

## 2. Materials and method

### 2.1. Data acquisition

The videos were collected on the Medau pig research and teaching farm (VetFarm, Pottenstein, Austria) of the University of Veterinary Medicine, Vienna, Austria. The experiment protocol was approved 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). Thirteen Austrian Large White sows and Landrace × Large White crossbred sows were included in the experiment. Six sows were kept in SWAP pens with an area of 6.0 $m^2$ (Jyden Bur A/S, Vemb, Denmark), two sows in Trapezoid pens with an area of 5.5 $m^2$ (Schauer Agrotronic GmbH, Prambachkirchen, Austria), and five sows in Wing pens with an area of 5.5 $m^2$ (Stewa Steinhuber GmbH, Sattledt, Austria). The sows were moved into the pens approximately 5 days before the expected date of farrowing. The videos were collected from when the sows were moved into the farrowing pens until weaning of piglets at four weeks of age. Only the video when the sows were not confined in crates were selected in this study. An IP camera (GV-BX 1300-KV, Geovision, Taipei, China) was placed above each pen at the height of 3 m to the ground, giving an overhead view of the whole pen. Additionally, infrared spotlights (IR-LED294S-90, Microlight, Bad Nauheim, Germany) were installed in order to allow night recording. The resolution of the videos was 1280 × 720 pixels and the frame rate was 30 fps. The videos collected after farrowing included piglets (average litter size: 11.64 ± 3.56), whereas the videos collected before farrowing without piglets. The piglets-included videos were used to test the robustness of the developed algorithm. Fig. 1 shows an example of frames including and not including piglets.

Recordings were stored on exchangeable, external 3 TB and 4 TB hard drives. The computer used in analysing the videos has a processor Intel(R) Core(TM) i7-8700 K CPU @ 3.70 GHz with 8 GB of RAM memory running a Microsoft Windows 10 Enterprise operating system. The graphic card was NVIDIA GeForce RTX 2080 with 8 GB of physical memory.

(a) (b)

**Fig. 1.** Examples of (a) a frame not including piglets; and (b) a frame including piglets.

### 2.2. Data sets

According to the definitions for sow's postural behaviour and the frequency of postural changes in the literature (Johnson et al., 2007; Oczak et al., 2016), a duration of 5 s was chosen for each video in this study. Only one behaviour was included in each video. The videos (including the training and test set) of each behaviour (sternal lying, lateral lying, sitting, standing, and walking) were labelled by one trained expert with extensive experience in sow's behaviour classification who was tested for inter-observer reliability. Different frames for each behaviour can be seen in Fig. 2, and the number of videos for each behaviour can be found in Table 1. For all the videos selected for this

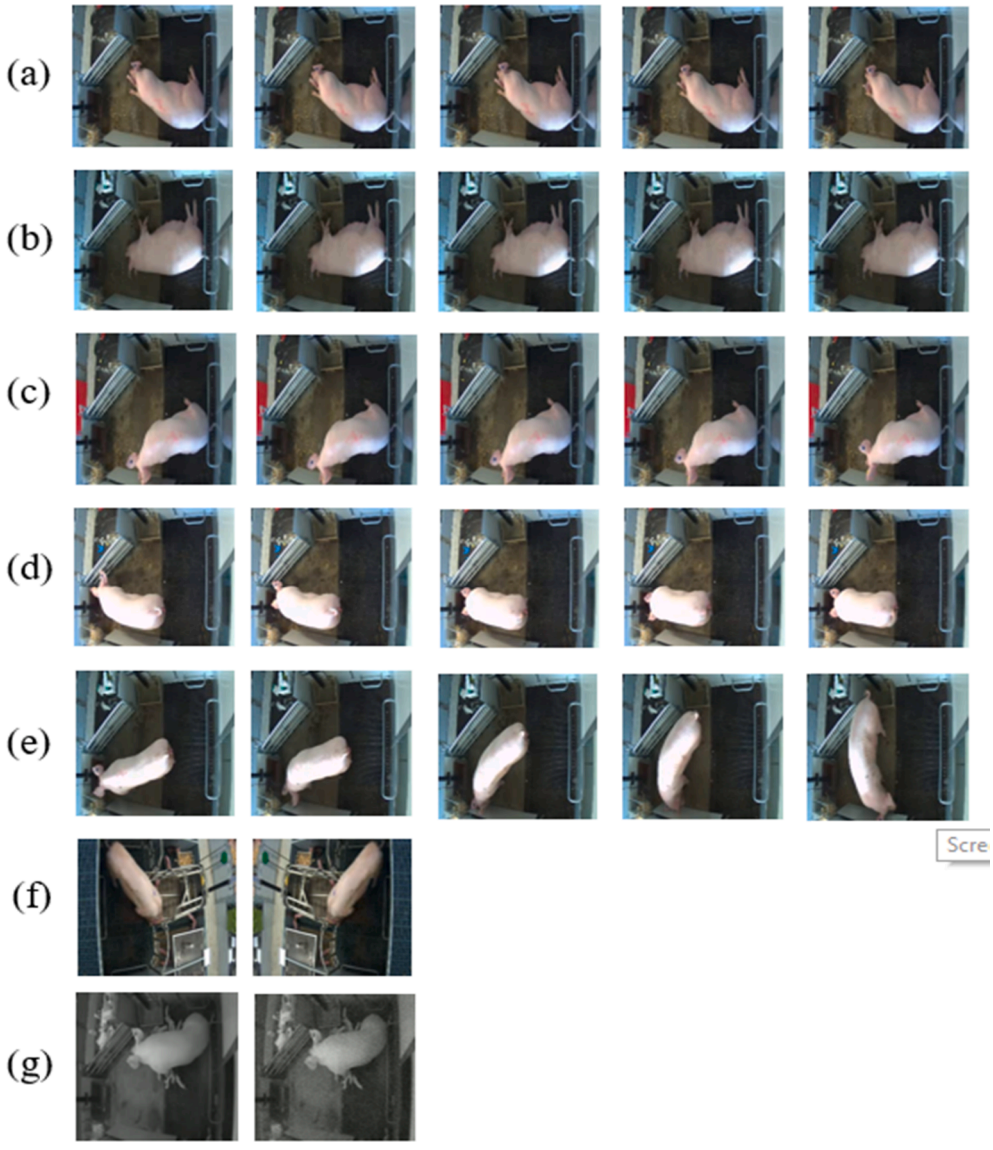

**Fig. 2.** Examples of frame sequences for different behaviours and the augmentation of vertical mirroring and adding Gaussian noise: (a) sternal lying; (b) lateral lying; (c) sitting; (d) standing; (e) walking; (f) vertical mirroring; (g) adding Gaussian noise (the right one was added Gaussian noise, which is less clear than left one.)

**Table 1**
Number of 5 s videos for each behaviour.

| | Sternal lying | Lateral lying | Sitting | Standing | Walking | Total |
|---|---|---|---|---|---|---|
| **Original videos without piglets** | 700 | 800 | 400 | 800 | 400 | 3100 |
| **Augmented videos without piglets** | 700 | 800 | 900 | 800 | 900 | 4100 |
| **Original videos with piglets** | 330 | 330 | 300 | 500 | 220 | 1680 |
| **Augmented videos with piglets** | 660 | 660 | 600 | 500 | 500 | 2920 |
| **Test set** | 100 | 100 | 100 | 100 | 100 | 500 |

study, 80% were collected during the day and the remaining 20% came from the night. In order to enhance the generalization ability and robustness of the model, an augmentation process was performed to the original data. The original video was augmented by adopting vertical mirror and adding random Gaussian noise. An illustration of the augmentation can be found in Fig. 2. The number of videos obtained after the augmentation process is shown in Table 1. Note that in the dataset without piglets, half of the videos of sternal lying, lateral lying and standing were augmented by applying adding Gaussian noise and half were added Gaussian noise respectively, and all the videos of sitting and walking were augmented by applying vertical mirror and adding Gaussian noise. In the dataset including piglets, half of the videos of standing were augmented by applying vertical mirror and half were added Gaussian noise, and all the videos of sternal lying, lateral lying, sitting and walking were augmented by applying vertical mirror and adding Gaussian noise. In order to test the robustness of the algorithm, there were two trainings: one (Training 1) on original and augmented videos without piglets (7200 videos in total), and the other one (Training 2) on all original and augmented data (11800 videos in total). Details for the number of videos used in Training 1 and 2 can be found in Fig. 3. In each training session, 75% of the data were randomly allocated as training set and the remaining 25% as validation set. Additionally, the trained models of Training 2 were tested on unseen videos that were collected on the same sows but were not used for training. Details for the number of videos for each behaviour in the test set are given in Table 1 and the number of videos used in test can be found in Fig. 3.

### 2.3. Algorithm

The workflow of the algorithm can be found in Fig. 4. The proposed algorithm is composed of two parts: (1) processing each video and (2) using a CNN + LSTM model to classify the behaviours. In the first part, a PCA was performed to reduce the dimension of the frames and then the frame distances were computed and compared with the threshold. Finally, the frames that satisfy the threshold requirement were selected to generate a new video for classification. Note that the threshold was different from videos as it was determined by the largest distance of two successive frames in each video. After the first part, only the motion-related frames were selected to generate the new videos. In the second part, the newly generated videos were classified by a CNN + LSTM model. First, a pre-trained CNN model was used to transform the frames into a feature vector. Then, the feature vectors were input to a LSTM module including the fully connected layer and *softmax* to extract the temporal features and to classify the different postural behaviours.

#### 2.3.1. Video processing

The cameras switched to IR mode when there was not enough light in the room, and in this mode videos were only recorded in grayscale. Due to the day-night difference in light intensity, the videos collected were almost half grayscale and half RGB format. Note that the raw videos in grayscale still have three channels but the intensity of the three channels are same. In order to speed up frame selection process, the algorithm converted the three channels into one channel at the beginning of video processing. Specifically, the means of the red, green and blue channels were computed. Then, each frame was represented by a 255 × 255-dimensional matrix, which is the dimension for training input of the CNN based on VGG16. Note that the one-channel data were only used with the aim to get motion-related frame indexes. After getting the indexes, the selection of the frames was still based on the raw video.

In this study, PCA was adopted to reduce the dimension of the frame matrix so that each frame could be represented in a lower dimension. Before doing PCA, the matrix (255 × 255 dimensions) that contains the frame data was flattened into a 65,025 (=255 × 255)-dimensional vector. By flattening each frame in the video, the video was finally represented by a 150 × 65025-dimensional matrix (5 sec video and 30 fps = 150 frames per video). Then, PCA was performed in this video matrix. The first k components were selected, and the video matrix was reduced to 150 × k dimensions. After the dimension reduction, each frame was represented by k components. The frame distances were computed based on these components. At first, the Euclidean distance of every pair of successive frames was calculated. Then, part of the maximum of these distances (*e.g.* one half) was set as the threshold. After that, the sum of the distances was computed from the first distance (between the first and second frames). If the sum of the preceding

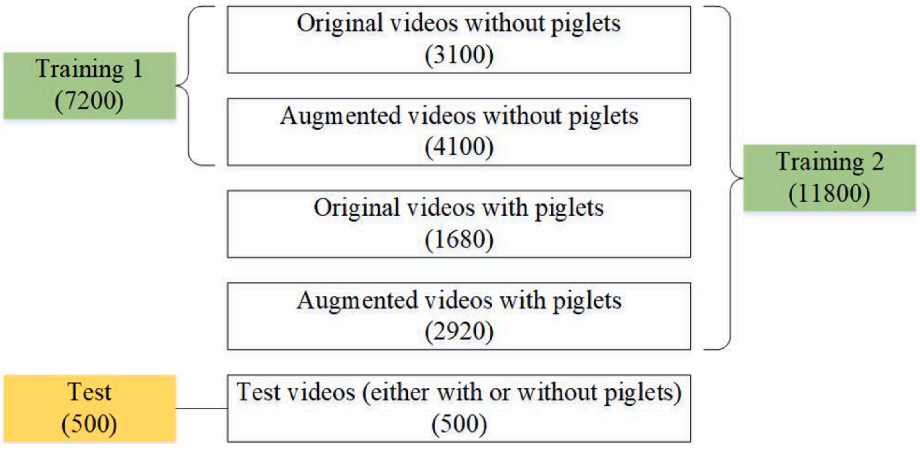


**Fig. 3.** Number of videos used for Training 1, 2 and Test.

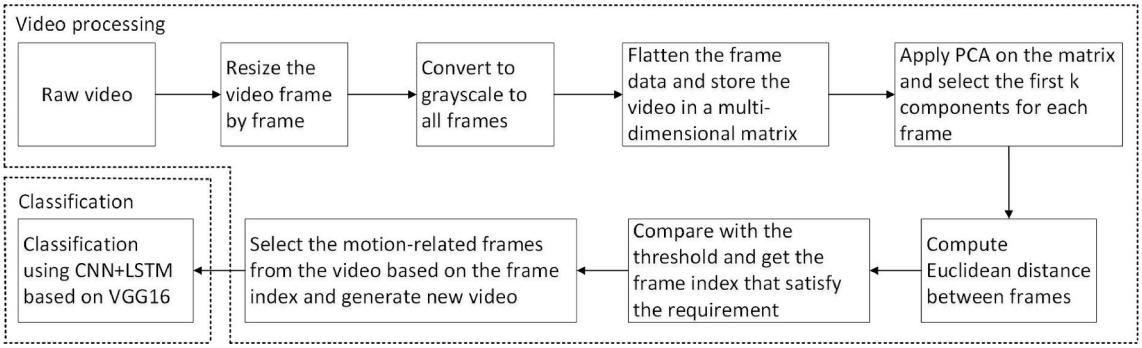


**Fig. 4.** Flow chart representation of the different processing steps for classifying different postural behaviours in sows.

distances exceeded the threshold, then that frame would be selected to generate the new video. This process was repeated until the last distance. The details of the whole computation of video processing is given below. The illustration of the frame selection can be found in Fig. 5.

**Algorithms 1**. (*Computation details of frame selection*)

**Assume:** Dataset V contains n videos $v_1, v_2, \cdots v_n$
Each video contains r frames, for example, the frames in video $v_i$ are $v_i = \{f_{i1}, f_{i2}, \cdots f_{ir}\}$

(*continued on next page*)

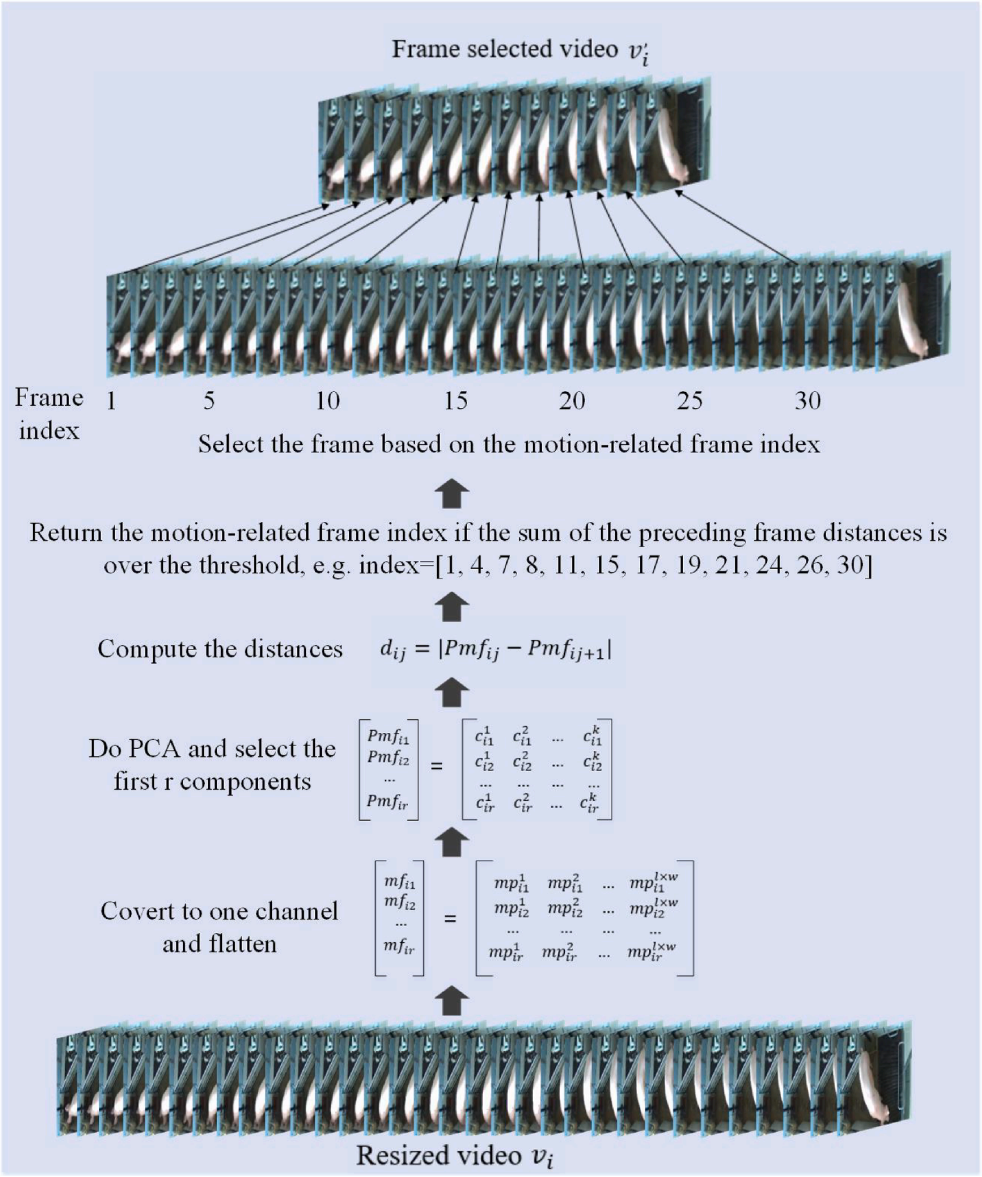


**Fig. 5.** The schematic diagram of processing a single video (5 s, 150 frames).

(*continued*)

The pixels in $f_{ij}$ are $f_{ij} = \{p_{ij}^{1,1}, p_{ij}^{1,2}, \cdots, p_{ij}^{l,w}\}$ , here l and w are the height and width of $f_{ij}$

The R, G, B values of each pixel are $p_{ij}^{a,b} = \{r_{ij}^{a,b}, g_{ij}^{a,b}, b_{ij}^{a,b}\}$, here a $\in \{1, 2, \cdots, l\}$,b $\in \{1, 2, \cdots, w\}$

Threshold T

Output dataset $V^{'} = \{v_1^{'}, v_2^{'}, \cdots v_n^{'}\}$

**Input:** Video sequences $V = \{v_1, v_2, \cdots v_n\}$

1: For $v_i \in$ V(i $= 1, 2, \cdots$n)

2: For $f_{ij} \in v_i$

3: For $p_{ij}^{a,b} \in f_{ij}$

4: $mp_{ij}^{a,b} = (r_{ij}^{a,b} + g_{ij}^{a,b} + b_{ij}^{a,b})/3$

5: End for

6: Flatten all pixels: $mf_{ij} = \{mp_{ij}^1, mp_{ij}^2, \cdots, mp_{ij}^{l\times w}\}$

7: End for

8: *let* $mv_i = \left\{mf_{ij}\right\}, i \in \{1, 2, \cdots, n\}, j \in \{1, 2, \cdots, r\}$ , the row and column numbers of $mv_i$ are $r$ *and* $l \times w$ respectively

9: Do PCA in $mv_i$ and select the first k components, $Pmv_i = PCA(mv_i) = \{Pmf_{ij}\}$, $Pmf_{ij} = \{c_{ij}^1, c_{ij}^2, \cdots, c_{ij}^k\}$

10: For $j \in \{1, 2, \cdots, r-1\}$

11: $d_{ij} = \left|Pmf_{ij} - Pmf_{ij+1}\right| = \sum_{s=1}^{k} \sqrt[2]{(c_{ij}^s - c_{ij+1}^s)^2}$

12: End for

13: Let $d = 0$

14: For $j \in \{1, 2, \cdots, k-1\}$

15: *if* $d \geq T$

16: Append $j + 1$to index, let $d = 0$

17: else

18: $d = d + d_{ij}$

19: *if*$d \geq T$

20: Append $j + 1$to index, let $d = 0$

21: End for

22: $v_i^{'} = v_i[index]$

**Output:** Frame selected video sequence $V^{'}$

### 2.3.2. Classification

In this study, a pre-trained CNN model VGG-16 (Simonyan and Zisserman, 2014.) was used to extract the spatial features. VGG-16 was trained on ImageNet (Deng et al., 2009), which has more than 14 million images and covers over 20,000 categories for object detection and classification. To fine-tune the model, the weights from the first to the penultimate layer were kept and the last layer's weights were trained on the frame-selected videos. Note that after selecting motion-related frames, the length of the videos were different (1.15 ± 0.24 s for videos not including piglets and 2.03 ± 0.25 s for videos including piglets). The input of the VGG-16 model were the frame-selected videos and the frame resolution was 224 × 224 × 3 pixels. VGG-16 extracted the spatial features frame by frame; the output of each frame was a 25088-dimensional vector. Then this vector was input to the LSTM module. LSTM can convey the features from the previous to the following frames. This makes LSTM have the function of memory, and thus able to extract temporal features. Fig. 6 shows the feature extraction process where first the VGG-16 CNN model was fine-tuned and then LSTM was applied to extract the temporal features. The output of the LSTM module was a 5-dimensional vector, i.e. one-hot encoding of the five different behaviours. Besides, categorical cross-entropy was used as the loss function when training the model.

Moreover, the accuracy defined in Eq.(1) was used to evaluate the proposed model. Here the number of accurately classified videos included all five behaviours that were classified correctly.

$$Acc = \frac{Number\ \ of\ \ accurately\ \ classified\ \ videos}{Total\ \ number\ \ of\ \ videos} \times 100\% \quad (1)$$

## 3. Results and discussion

### 3.1. Results of training 1 and training 2

Fig. 7(a) illustrates the first 30 frames in a video and Fig. 7(b) shows that 6 frames were selected out of the 30 frames. In the part of video processing, different numbers of components were used to represent the frame when performing PCA to reduce the dimensions. The number of components can affect the number of selected frames and thus change the accuracy and duration of the training. Table 2 illustrates the training time and accuracy under different number of components in Training 1. The PCA time and Training time in Table 2 indicate the time used for doing PCA and training respectively. The average frame number of each video and the percentage of variance kept in Training 1 are reported in Table 2. The similar parameters of Training 2 are reported in Table 3. Note that in Table 3 'All training time' was the time used including doing PCA and training. Additionally, Table 3 gives the processing time for using the trained model on all data (including and not including piglets, 11,800 videos in total). By computing the processing time we want to show the model efficiency when we use the pre-trained model to do classification. Note that the distance thresholds here were all set to M/2, where M indicates the maximum distance between two successive frames. M was different for each video.

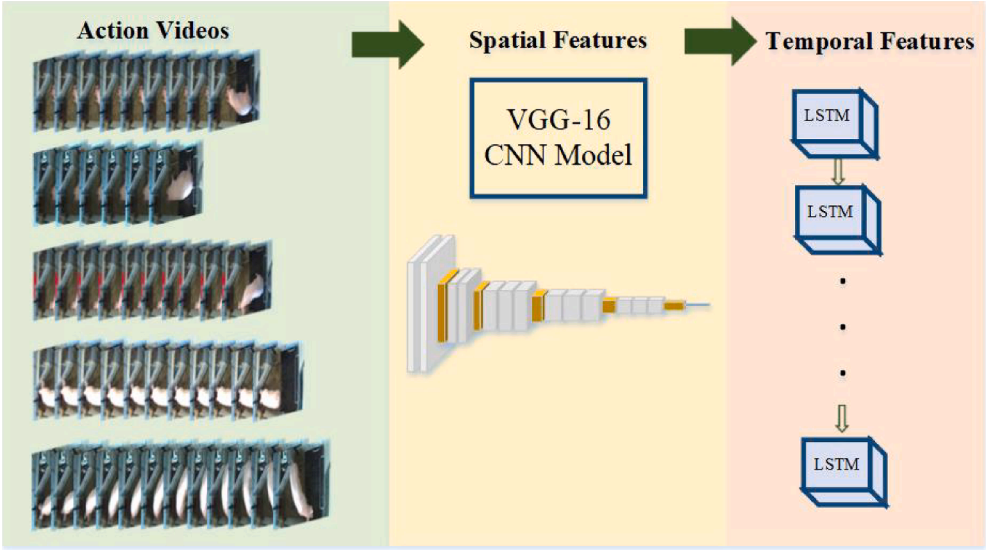


**Fig. 6.** Illustration of spatial and temporal features extraction.


**Fig. 7.** Illustration of before and after PCA-based frame selection: (a) before frame selection; (b) after frame selection.

**Table 2**
Results of Training 1.

| Number of components | 1 | 2 | 3 | 5 | 10 | 20 | 50 | Raw video |
|---|---|---|---|---|---|---|---|---|
| **Training Accuracy** | 94.09% | 94.36% | 94.79% | 94.90% | 94.93% | 95.04% | 95.69% | 97.33% |
| **Validation Accuracy** | 93.11% | 93.19% | 93.39% | 93.51% | 93.58% | 93.58% | 94.34% | 95.97% |
| **Percent of variance kept** | 38.51% | 57.72% | 67.50% | 77.91% | 87.98% | 94.40% | 98.39% | 100% |
| **PCA time** | 13.48 s | 13.35 s | 14.53 s | 14.47 s | 15.37 s | 15.48 s | 16.22 s | / |
| **Training time** | 10m43s | 12m05s | 12m44s | 12m28s | 12m51s | 13m26s | 13m47s | 18m55s |
| **Average frame number per video** | 27.01 | 29.87 | 30.42 | 31.86 | 32.95 | 38.33 | 50.48 | 150 |

**Table 3**
Results of Training 2.

| Number of components | 1 | 2 | 3 | 5 | 10 | 20 | 50 | Raw video |
|---|---|---|---|---|---|---|---|---|
| **Training Accuracy** | 91.35% | 91.50% | 93.35% | 93.87% | 93.75% | 93.77% | 93.81% | 95.81% |
| **Validation Accuracy** | 88.01% | 88.50% | 88.55% | 91.48% | 92.67% | 92.74% | 92.85% | 95.68% |
| **Percent of variance kept** | 38.51% | 57.72% | 67.50% | 77.91% | 87.98% | 94.40% | 98.39% | 100% |
| **All training time** | 18m25s | 18m17s | 18m34s | 18m31s | 19m02s | 20m12s | 21m09s | 28m44s |
| **Processing time** | 8m19s | 8m22s | 9m47s | 8m38s | 9m03s | 9m32s | 10m40s | 26m21s |
| **Average frame number per video** | 50.17 | 54.34 | 57.69 | 60.83 | 61.68 | 65.75 | 74.81 | 150 |

From Table 2 we can see that the more components were used, the more frames were selected, and larger variance were kept compared to the raw videos. Both the accuracy and training time increased with increasing numbers of components. Compared to using raw videos, only using motion-related frames saved about one third of training time while the performance were almost retained. Similar conclusions can also be drawn from Table 3. From Table 2 to Table 3, it can be seen that the algorithm was also able to analyze data including piglets. It should be noticed that in Training 1 the validation accuracy increased by 0.78% when the number of components increased from 10 to 50, but the training time also increased by 2m35s. More importantly, the validation accuracy of data including piglets in Training 2 only improved by 0.20% with more than 10 components while the training time increased 2m07s. Considering the trade-off between training time and validation accuracy, we chose 10 components as an optimal representation of the frame for classifying different behaviours, and thus this number was used in the following tests. Fig. 8 (a) and (c) illustrate the training curve when using 10 components in Training 1 and 2. Additionally, in order to verify the model efficiency, the trained models were all validated on the whole dataset (including and not including piglets) and the processing time were recorded in Table 3. We can see that using motion-based frames can save over half of the time compared to using all frames in the video.

### 3.2. *Testing different thresholds*

In the previous steps, the threshold was set as M/2. However, the threshold can also affect the number of selected frames. To verify how this parameter affects the results, different thresholds were tested on dataset of Training 1 and the results are illustrated in Table 4, with the number of components set to 10 as mentioned above. Table 4 reveals that by setting a larger threshold, a smaller number of frames will be selected. Therefore, it costs less time to train the model. From the training and validation accuracy, we can see that selecting more frames did not necessarily result in higher accuracy. The accuracies obtained by setting the threshold to M/4 were better than those of M/6 and M/8. This may result from the selection of irrelevant frames and these frames may cause the misclassification of the behaviours. Overall, the results obtained from threshold M/4 were optimal considering both the training time and accuracy, and therefore it was used for further tests. Fig. 8 (b) illustrates the training curve of setting the threshold to M/4 and using 10 components. By comparing Fig. 8 (a) and (b), it can be seen that by setting a suitable threshold, the validation accuracy increased from 93.58% to 95.33% and training time only increased by 20 s.

### 3.3. *Comparison with cluster-based method*

In order to validate the effectiveness, we made the comparison with cluster-based method. The details of the algorithm can be found in

**Table 4**
Results of testing different thresholds on the dataset of Training 1.

| Threshold | M/2 | M/3 | M/4 | M/5 | M/6 | M/8 | M/10 | Raw video |
|---|---|---|---|---|---|---|---|---|
| **Training Accuracy** | 94.91% | 95.83% | 96.23% | 95.69% | 95.70% | 96.57% | 96.33% | 97.35% |
| **Validation Accuracy** | 93.58% | 95.19% | 95.33% | 94.83% | 94.95% | 95.91% | 95.91% | 95.99% |
| **All Training time** | 13m06s | 13m24s | 13m26s | 13m45s | 14m56s | 15m01s | 15m05s | 19m16s |
| **Average frame number per video** | 38.33 | 51.86 | 61.78 | 68.47 | 73.91 | 82.41 | 88.93 | 150 |

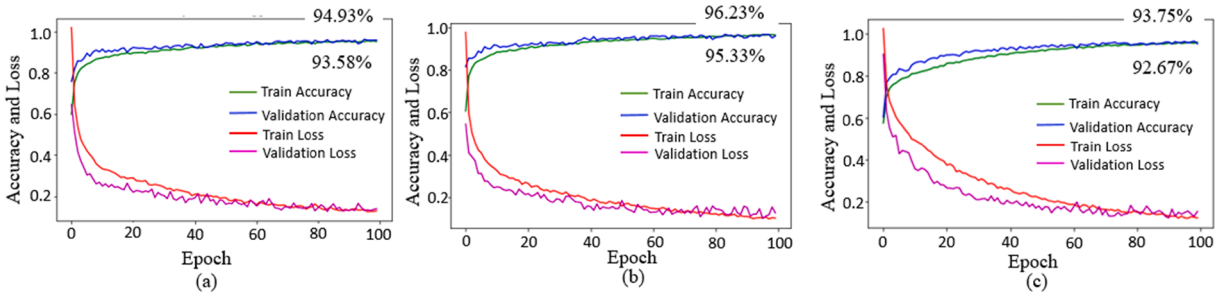


**Fig. 8.** Training curves: (a) Training 1, number of components = 10, threshold = M/2; (b) Training 1, number of components = 10, threshold = M/4; (c) Training 2, number of components = 10, threshold = M/2;

(Zhuang et al., 1998). We tested different threshold parameter δ (same as the reference, $\delta = 0.8$, 0.85 and 0.9), which controls the number of selected frames. The clustering time, training time, average frame per video as well as validation accuracy are showed in Table 5. Note that the validation was performed on the dataset of training 1. From Table 5 we can see that cluster-based method can exclude a large number of frames so the training time decreased a lot. As cluster-based method computer the cluster centroids iteratively, the clustering time was longer than the time of performing PCA. The greatest difference of these two methods was the validation accuracy, we can see that the validation accuracy of cluster-based method was around 80%, which was low for classification. The cluster-based method was a good tool for video summarization since only a few frames were kept. But only a few frames seems not enough for postural behaviour classification. We believe there is a trade-off between the frame number and classification accuracy, PCA-based method might be more applicable to less complex scenario, *e.g.* postural behaviour classification in sows.

### 3.4. Testing on unseen data

After confirming the best component number and threshold, the final model was trained on the dataset of Training 2 with 10 components and M/4 threshold. In order to see the generalization ability of the model, unseen new data were used to test the performance of the model. The number of videos for each behaviour in the new test dataset was 100 with 500 videos in total. The new test dataset included both videos with and without piglets, from which 40% were data including piglets. The average accuracy obtained was 90.60%. It can be seen that the proposed method is able to be generalized on new datasets. The confusion matrix of the test is shown in Table 6. Analysing the confusion matrix and the videos, we found that the misclassification mainly resulted from the overlaps between the sow and piglets. and also the movements of piglets affecting the sow's behaviour especially in corners of the image. Fig. 9 illustrates the possible reasons for the misclassification between standing and walking.

(1) In Fig. 9(a), the piglet moved around the rear of the sow and finally vanished and overlapped with the sow, which may have been considered as the sow's body lifting in the air. Therefore, this behaviour was falsely classified as standing.
(2) In Fig. 9(b), the piglets moved around the sow's hind and the sow was in the corner, so that the sow's hinds tucked in the corner looked like the hinds touching the floor. Thus, this behaviour was falsely classified as sitting.

**Table 5**
Comparison with cluster-based methods.

| | PCA | $\delta = 0.8$ | $\delta = 0.85$ | $\delta = 0.9$ |
|---|---|---|---|---|
| **Average frame number per video** | 61.78 | 3.00 | 5.86 | 9.47 |
| **PCA/clustering time** | 15.46 s | 25.28 s | 24.31 s | 25.66 s |
| **Training time** | 12m38s | 4m19s | 5m07s | 5m48s |
| **Validation accuracy** | 94.83% | 72.18% | 80.40% | 77.35% |

### 3.5. Discussion and future work

In previous research on behaviour classification based on video analysis, the main objective was to classify different behaviours as accurately as possible (Zhang et al., 2020; Li et al., 2020). The basis for classification of postural behaviours was the extraction of temporal and spatial features from videos. However, previous studies all extracted these features from the whole video. In research by Zhang *et al.* (Zhang et al., 2020), who used a Two-Stream Convolutional Networks for behaviour classification, the average processing time of each video was 0.3163 s. It was potential to improve this result by 85% by selecting motion-related frames given that the processing time per video in our study was 0.0462 s (543/11,800 = 0.0462 s). Furthermore, the number of selected frames were different for different behaviours. Table 7 shows the average frame numbers per video for different behaviours in Training 2. It is obvious that the frame numbers of lying behaviours are less than walking and standing, which indicates that the processing time for lying behaviours would save more time than for other behaviours. On the other hand, sows spent about 60–70% of their time lying during the night (Teng and Yu, 2017). Thus, in this case, the proposed frame selection method can help save more time at night.

In this study, the data without piglets included the period covering approximately 5 days before the sow's expected date of farrowing. Around the time of farrowing, sows are still commonly confined in a farrowing crate in order to protect the piglets from being crushed by the sow after the start of farrowing, but this is a compromise that impairs the sow's welfare to the benefit of her piglets and the farmer (King et al., 2019). This impairment currently affects all sows, regardless of whether they actually crush piglets or not. Recently, temporary crating has been introduced to loose housing of farrowing and lactating sows. According to the concept of temporary crating, sows could be kept out of the crate

**Table 6**
The confusion matrix of testing on unseen dataset.

| | Sternal lying | Lateral lying | Sitting | Standing | Walking | Total |
|---|---|---|---|---|---|---|
| **Sternal lying** | 98 | 0 | 2 | 0 | 0 | 100 |
| **Lateral lying** | 1 | 99 | 0 | 0 | 0 | 100 |
| **Sitting** | 5 | 0 | 80 | 10 | 5 | 100 |
| **Standing** | 6 | 0 | 10 | 83 | 1 | 100 |
| **Walking** | 0 | 0 | 0 | 7 | 93 | 100 |
| **Total** | 109 | 100 | 98 | 97 | 96 | 500 |

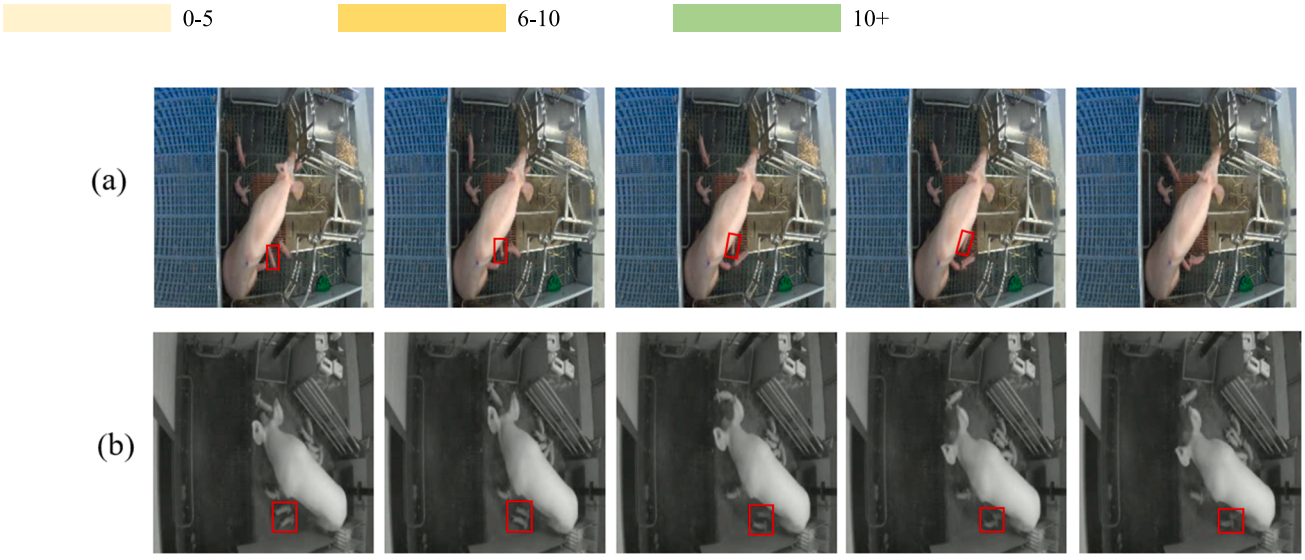


**Fig. 9.** Examples for misclassification: (a) sitting falsely classifies as standing; (b) standing falsely classified as sitting.

**Table 7**
The average frame numbers per video of different behaviours in Training 2.

| Number of components | 1 | 2 | 3 | 5 | 10 | 20 | 50 | Raw video |
|---|---|---|---|---|---|---|---|---|
| **Lateral lying** | 33.10 | 35.00 | 35.84 | 36.23 | 36.90 | 45.25 | 62.96 | 150 |
| **Sternal lying** | 49.11 | 56.22 | 58.38 | 59.45 | 56.10 | 59.75 | 71.42 | 150 |
| **Sitting** | 51.44 | 56.18 | 60.78 | 63.92 | 63.95 | 65.77 | 74.51 | 150 |
| **Standing** | 61.00 | 62.37 | 66.96 | 70.93 | 71.79 | 74.46 | 78.80 | 150 |
| **Walking** | 57.45 | 63.99 | 69.26 | 77.91 | 85.94 | 89.46 | 90.96 | 150 |

at the time of prenatal nest-building behaviour and after the critical period of piglets' life (Oczak et al., 2020). It is therefore important for the farmer to find the right time for confining the sow in the crate after the end of nest-building and before the beginning of farrowing to safeguard both sow and piglet welfare. Previous studies indicated that the basis for estimation of onset time of farrowing was the increase of activity related to nest-building behaviour (Castrén et al., 1993; Erez and Hartsock, 1990). Recent studies mainly used accelerometers to estimate the onset time of farrowing (Oczak et al., 2020; Pastell et al., 2016; Traulsen et al., 2018). Although effective, contactless techniques based on video analysis will likely be favoured in pig farming due to their non-invasive and practical characteristics. To monitor the increased activity by using video analysis, the frequency of each postural behaviour could be accurately and efficiently calculated. The proposed model could first be applied to videos of the sow before farrowing to classify different behaviours and then the frequency of each behaviour can be computed based on the output of classification. Finally, a threshold related to activity frequency should be set to confirm the onset of farrowing.

In addition to the videos captured before farrowing, this study also included videos after the piglets were born, which on one hand helped test the robustness of the algorithm, and on the other hand might be useful in assessing the risk of piglet crushing, based on postural changes (Nicolaisen et al., 2019; Damm et al., 2005) as well as potentially detecting piglet crushing events. The most common mean to prevent piglet crushing in indoor farming systems remains the use of farrowing crate (Weber et al., 2007). However, this restriction of movement has negative effect on sow welfare (King et al., 2019). Actively prevent piglet crushing, in contrast to the passive principle of the farrowing crate, is less common in practice. Experienced personnel are often able to recognize a trapped piglet acoustically. In such a case, a behavioural change of the mother sow would be stimulated by manually forcing her up. A survival rate of about 95% was found for piglets trapped less than 1 m (Weary et al., 1996). After being trapped for up to 4 m, still about 33% of the piglets would survive (Weary et al., 1996). Therefore, the survival of the crushed piglets depends on the quick behaviour changes of the sow, and if the behaviour changes of the sow can be monitored after the crushing it is able to know if the piglet is set free or not. Regarding the detection of piglet crushing, presently it was mainly realized by vocalization analysis (Manteuffel et al., 2017; Chen et al., 2019). Although the detection was reliable, the follow-up behaviours of the sow were lost. Replay studies found 60% (Hutson et al., 1991) to 100% (Cronin and Cropley, 1991) of sows reacting when a piglet dummy was placed underneath their body and piglet squeals were played, which indicates that almost all the crushing can be avoided by the sow so that the human handling is only needed in limited cases. Therefore, there has been great interest in monitoring the follow-up behaviour of the sow

after crushing event by using video analysis. Considering the proposed method already realized the behaviour classification of the sow while the piglets were present, the next step for this task is to detect the crushing event, which should also include the position detection and tracking of the piglet.

## 4. Conclusion

A PCA-based frame selection method was developed to classify sow's postural behaviours using deep learning. The videos including piglets were used to test the robustness of the method. The accuracies reached 95.33% and 92.67% on videos without piglets and all video data (including and not including piglets), respectively. Additionally, the computation time was reduced by about one third by scarifying only 0.7% of accuracy compared to using raw videos. By testing different component numbers and thresholds, the best results were achieved with using 10 components and setting the threshold to one fourth of the largest distance between two successive frames. Using these optimised parameters, the accuracy of the test on unseen data reached 90.60%, indicating that the algorithm can be generalized to new data. Future work can be focused on applying this method to improve animal welfare, *e.g.* monitoring sow's follow-up behaviours after piglet crushing.

## CRediT authorship contribution statement

**Meiqing Wang:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Validation, Visualization, Writing – original draft, Writing - review & editing. **Maciek oczak:** Data curation, Project administration, Resources, Writing - review & editing. **Mona Larsen:** Supervision, Writing - review & editing. **Bayer Florian:** Data curation, Writing - review & editing. **Kristina Maschat:** Data curation, Writing - review & editing. **Johannes Baumgartner:** Data curation, Writing - review & editing. **Jean-Loup Rault:** Project administration, Resources, Supervision, Writing - review & editing. **Tomas Norton:** Supervision, Writing - review & editing.

## 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.

## References


Bonde, M., Rousing, T., Badsberg, J.H., Sørensen, J.T., 2004. Associations between lying-down behaviour problems and body condition, limb disorders and skin lesions of lactating sows housed in farrowing crates in commercial sow herds. Livest. Prod. Sci. 87, 179–187.

Castrén, H., Algers, B., De Passille, 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, 91–102.

Chen, W.-E., Chen, L.-X., Chiu, Y.-C., 2019. An Intelligent Detection and Notification (iDN) System for Handling Piglet Crushing Based on Machine Learning. In: Proceedings of the International Cognitive Cities Conference. Springer, pp. 475–484.

Chen, C., Zhu, W., Steibel, J., Siegford, J., Han, J., Norton, T., 2020. Classification of drinking and drinker-playing in pigs by a video-based deep learning method. Biosyst. Eng. 196, 1–14.

Chen, C., Zhu, W., Steibel, J., Siegford, J., Wurtz, K., Han, J., Norton, T., 2020. Recognition of aggressive episodes of pigs based on convolutional neural network and long short-term memory. Comput. Electron. Agric. 169, 105166.

Cronin, G.M., Cropley, J.A., 1991. The effect of piglet stimuli on the posture changing behaviour of recently farrowed sows. Appl. Anim. Behav. Sci. 30, 167–172.

Damm, B.I., Forkman, B., Pedersen, L.J., 2005. Lying down and rolling behaviour in sows in relation to piglet crushing. Appl. Anim. Behav. Sci. 90, 3–20.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Imagenet, F.-F., 2009. A large-scale hierarchical image database. In: Proceedings of the 2009 IEEE conference on computer vision and pattern recognition, pp. 248–255.

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.

Ferman, A.M., Tekalp, A.M., 1997. Multiscale content extraction and representation for video indexing. In: Proceedings of the Multimedia Storage and Archiving Systems II. International Society for Optics and Photonics, vol. 3229, pp. 23–31.

Grégoire, J., Bergeron, R., d'Allaire, S., Meunier-Salaün, M.-C., Devillers, N., 2013. Assessment of lameness in sows using gait, footprints, postural behaviour and foot lesion analysis. Anim. an Int. J. Anim. Biosci. 7, 1163.

Hutson, G.D., Wilkinson, J.L., Luxford, B.G., 1991. The response of lactating sows to tactile, visual and auditory stimuli associated with a model piglet. Appl. Anim. Behav. Sci. 32, 129–137.

Huynh, T.T.T., Aarnink, A.J.A., Gerrits, W.J.J., Heetkamp, M.J.H., Canh, T.T., Spoolder, H.A.M., Kemp, B., Verstegen, M.W.A., 2005. Thermal behaviour of growing pigs in response to high temperature and humidity. Appl. Anim. Behav. Sci. 91, 1–16.

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, 59–74.

Kashiha, M.A., Bahr, C., Ott, S., Moons, C.P.H., Niewold, T.A., Tuyttens, F., Berckmans, D., 2013. Automatic Monitoring of Pig Activity Using Image Analysis. In: BlancTalon, J., Kasinski, A., Philips, W., Popescu, D., Scheunders, P., (Eds.). Proceedings of the ADVANCED CONCEPTS FOR INTELLIGENT VISION SYSTEMS, ACIVS 2013, vol. 8192, pp. 555–563.

King, R.L., Baxter, E.M., Matheson, S.M., Edwards, S.A., 2019. Temporary crate opening procedure affects immediate post-opening piglet mortality and sow behaviour. Animal 13, 189–197.

Larsen, T., Kaiser, M., Herskin, M.S., 2015. Does the presence of shoulder ulcers affect the behaviour of sows? Res. Vet. Sci. 98, 19–24.

Li, D., Zhang, K., Li, Z., Chen, Y., 2020. A Spatiotemporal Convolutional Network for Multi-Behavior Recognition of Pigs. Sensors 20 (8), 2381. https://doi.org/10.3390/s20082381.

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.

Manteuffel, C., Hartung, E., Schmidt, M., Hoffmann, G., Schoen, P.C., 2017. Online detection and localisation of piglet crushing using vocalisation analysis and context data. Comput. Electron. Agric. 135, 108–114. https://doi.org/10.1016/j.compag.2016.12.017.

Marchant, J.N., Broom, D.M., Corning, S., 2001. The influence of sow behaviour on piglet mortality due to crushing in an open farrowing system. Anim. Sci. 72, 19–28.

Matai, J., Kastner, R., Cutter, G.R., Demer, D.A., 2012. Automated techniques for detection and recognition of fishes using computer vision algorithms. In: Proceedings of the Report of the National Marine Fisheries Service Automated Image Processing Workshop, pp. 35–37.

Nasirahmadi, A., Sturm, B., Olsson, A.-C., Jeppsson, K.-H., Mueller, S., Edwards, S., Hensel, O., 2019. Automatic scoring of lateral and sternal lying posture in grouped pigs using image processing and Support Vector Machine. Comput. Electron. Agric. 156, 475–481. https://doi.org/10.1016/j.compag.2018.12.009.

Nicolaisen, T., Lühken, E., Volkmann, N., Rohn, K., Kemper, N., Fels, M., 2019. The Effect of Sows' and Piglets' Behaviour on Piglet Crushing Patterns in Two Different Farrowing Pen Systems. Animals 9 (8), 538. https://doi.org/10.3390/ani9080538.

Oczak, M., Maschat, K., Berckmans, D., Vranken, E., Baumgartner, J., 2016. Can an automated labelling method based on accelerometer data replace a human labeller?–Postural profile of farrowing sows. Comput. Electron. Agric. 127, 168–175.

Oczak, M., Maschat, K., Baumgartner, J., 2020. 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.

Pastell, M., Flietaoja, J., Yun, J., Tiusanen, J., Valros, A., 2016. Predicting farrowing of sows housed in crates and pens using accelerometers and CUSUM charts. Comput. Electron. Agric. 127, 197–203. https://doi.org/10.1016/j.compag.2016.06.009.

Schindler, K., Van Gool, L., 2008. Action snippets: How many frames does human action recognition require?. In: Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–8.

Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. arXiv Prepr. arXiv1409.1556.

Spoolder, H.A.M., Aarnink, A.A.J., Vermeer, H.M., van Riel, J., Edwards, S.A., 2012. Effect of increasing temperature on space requirements of group housed finishing pigs. Appl. Anim. Behav. Sci. 138, 229–239.

Sun, L., Liu, Y., Chen, S., Luo, B., Li, Y., Liu, C., 2019. Pig Detection Algorithm Based on Sliding Windows and PCA Convolution. IEEE Access 7, 44229–44238.

Teng, G., Yu, Q., 2017. Pig behavior research and its application in breeding-landrace pigs as an example.

Traulsen, I., Scheel, C., Auer, W., Burfeind, O., Krieter, J., 2018. Using acceleration data to automatically detect the onset of farrowing in sows. Sensors 18 (2), 170. https://doi.org/10.3390/s18010170.

Weary, D.M., Pajor, E.A., Fraser, D., Honkanen, A.-M., 1996. Sow body movements that crush piglets: a comparison between two types of farrowing accommodation. Appl. Anim. Behav. Sci. 49, 149–158.

Weber, R., Keil, N.M., Fehr, M., Horat, R., 2007. Piglet mortality on farms using farrowing systems with or without crates. Anim. Welfare-Potters Bar Then Wheathampstead- 16, 277.

Yang, A., Huang, H., Zheng, C., Zhu, X., Yang, X., Chen, P., Xue, Y., 2018. High-accuracy image segmentation for lactating sows using a fully convolutional network. Biosyst. Eng. 176, 36–47.

Zhang, K., Li, D., Huang, J., Chen, Y., 2020. Automated Video Behavior Recognition of Pigs Using Two-Stream Convolutional Networks. Sensors 20 (4), 1085. https://doi.org/10.3390/s20041085.

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. https://doi.org/10.1016/j.compag.2018.01.023.

Zhuang, Y., Rui, Y., Huang, T.S., Mehrotra, S. (1998). Adaptive key frame extraction using unsupervised clustering. In: Proceedings of the Proceedings 1998 international conference on image processing. icip98 (cat. no. 98cb36269). IEEE, vol. 1, pp. 866–870.