HAL
open science

Predicting sow postures from video images: Comparison of convolutional neural networks and segmentation combined with support vector machines under various training and testing setups

Mathieu Bonneau, Bernard Benet, Yann Labrune, Jean Bailly, Edmond Ricard, Laurianne L. Canario

▶ **To cite this version:**

HAL Id: hal-03413048
https://hal.inrae.fr/hal-03413048

Submitted on 29 Jul 2022

Research Paper

Predicting sow postures from video images: Comparison of convolutional neural networks and segmentation combined with support vector machines under various training and testing setups

Mathieu Bonneau [a,*], Bernard Benet [b], Yann Labrune [c], Jean Bailly [d], Edmond Ricard [c], Laurianne Canario [c]

[a] URZ, INRA, 97170, Petit-Bourg (Guadeloupe), France
[b] UR TSCF, INRAE, Aubière, 63178, France
[c] UMR1388 GenPhySE, INRAE/INPT ENSAT/INPT ENVT, 31326, Castanet, Tolosan, France
[d] UE GenESI, INRAE, 17700, Surgères, France

ARTICLE INFO

The use of CNN and segmentation to extract image features for the prediction of four postures for sows kept in crates was examined. The extracted features were used as input variables in an SVM classification method to estimate posture. The possibility of using a posture prediction model with images not necessarily obtained under the same conditions as those used for the training set was explored. As a reference case, the efficacy of the posture prediction model was explored when training and testing datasets were built using the same pool of images. In this case, all the models produced satisfactory results, with a maximum f1-score of 97.7% with CNNs and 93.3% with segmentation. To evaluate the impact of environmental variations, the models were trained and tested on different monitoring days. In this case, the best f1-score dropped to 86.7%. The impact of using the posture prediction model on animals that were not present in the training dataset was then explored. The best f1-score reduced to 63.4% when the posture prediction models were trained on one animal and tested on 11 other different animals. Conversely, when the models were tested on one animal and trained on the 11 others, the f1-score only decreased to 86% with the best model. On average, a decrease of around 17% caused by environmental and individual variations between training and testing was observed.

Notations	
CNN	Convolutional neural network
SVM	Support vector machine
PPM	Posture prediction method
FEM	Feature extraction method
Seg-FEM	Feature extraction method based on segmentation
CNN-FEM	Feature extraction method based on CNN

1. Introduction

The survival of the offspring of farmed mammals is strongly influenced by the behaviour of their mother (Turner & Lawrence, 2007; Dwyer, 2014). Piglet mortality, which can be as high as 30%, is highly dependent on sow behaviour (e.g. Pedersen et al., 2006; Andersen et al., 2005; Canario et al., 2014; Nicolaisen et al., 2019). In most conventional herds, sows are kept in a crate during the lactation period, without being able to move or turn around. In line with societal concerns about animal welfare, farrowing systems are currently evolving towards looser housing conditions (CWF, 2021). The new housing conditions for sows (Baxter et al., 2018) are expected to become one of the main types of housing in a few years (Commision European, 2021). But their effect on piglet mortality has not yet been sufficiently investigated. To facilitate this transition, sow behaviour should be taken into account in breeding schemes to select sows with good maternal behaviour. In particular, sows that are calm around farrowing and careful when changing their posture have lower stillbirth and mortality rates due to crushing (Canario et al., 2014; Andersen et al., 2005). Comparing the level of activity between different types of housing can help assess the impact of their environment on animal health and welfare. Monitoring animals over a long periods of time may be necessary to detect changes in their activity level. However, only a few studies have attempted to assess the impact of sow housing on sow behaviour, and on an *ad hoc* basis during lactation. For example (Wallenbeck et al., 2008) analysed at 4 days post partum using a scan sampling methodology applied 24 h a day (Nicolaisen et al., 2019) analysed the first three days after parturition, also with a scan sampling methodology but using visual observations, which makes it almost impossible to create postural activity datasets that are uniform in the long term (Vizcarra & Wettemann, 1996; Nalon et al., 2014). In the last decade, a number of methods have been developed to automatically measure activity in pigs, mainly using accelerometers (Ringgenberg et al., 2010; Brown et al., 2013; Escalante et al., 2013; Oczak et al., 2016b; Matheson et al., 2017; Canario et al., 2019; Thompson et al., 2019; Chapa et al., 2020). Accelerometers are cheap, and the acquisition and analysis of the signal is relatively simple. However, in practice, the quality of the prediction depends on where the sensor is positioned on the animal, i.e. on the ear, neck, back or leg. Unfortunately, accelerometers can also be displaced during data acquisition due to movements by the animal itself or by other animals touching the sensor, which alters the prediction. More recently, computer vision analysis has emerged as a suitable tool for automatic monitoring of animal behaviour in several species (Aydin et al., 2010; Kashiha et al., 2013b; Li et al., 2017; Nasirahmadi et al., 2017; Bezen et al., 2020; Bonneau et al., 2020). Computer vision analysis consists in locating the animal in the image and then estimating several features. For example, in pigs, this can include live weight (Shi et al., 2016), animal identification (Marsot et al., 2020), the number of piglets (Oczak et al., 2016a), water consumption (Kashiha et al., 2013a), episodes of aggression (Viazzi et al., 2014; Chen et al., 2020) and different postures (Leonard et al., 2019; Nasirahmadi et al., 2019). Positioning the cameras on the housing disturbs animals less than embedded sensors. However, image analysis requires specific programming skills that are not necessarily available in research teams who study animal behaviour. We compared two common methods used to extract useful information, called features, from the image: image segmentation and convolutional neural networks (CNN). Once the features are extracted from the image, they can be used as input variables of a classification method, like those that use a support vector machine (SVM). Segmentation consists in applying image segmentation techniques to differentiate between the animal and background pixels. The extracted features could be geometric characteristics of the segmented image (Oczak et al., 2016a; Yang et al. 2018, 2019; Leonard et al., 2019; Nasirahmadi et al., 2019). The second method consists in using available CNNs for image analysis. The CNNs are already pretrained on a large image dataset, such as ImageNet (Deng et al., 2009), which generally contain more than a million images (Zhang et al., 2019; Chen et al., 2020; Marsot et al., 2020; Gan et al., 2021; Kasani et al., 2021). Both methods have been reported to produce promising results in the literature. However, even though software is available to aid the use of CNNs (e.g. Abadi et al., 2015; Moses & Olafenwa, 2018), it can still be difficult for new users. In this article, the use of the two methods to extract image features is compared and they are then used to estimate the postures of sows kept in a crate. Also investigated was the impact of the training dataset on the quality of the prediction model.

2. Materials and methods

2.1. Housing and imaging

This study was carried out using 12 Large White sows belonging to the INRAE experimental herd at GENESI Le Magneraud (Charente-Maritime, France. https://doi.org/10.15454/1.5572415481185847E12). Animals are raised in three-week batches, and two to three sows were measured in three successive batches between January and March 2020. The dimensions of the crate, and the location and dimensions of the cage inside the crate are given in our supplementary material S1. All the crates were equipped with a liquid feeding trough and one drinking nipple. The creep area for the piglets was equipped with artificial light that was kept on 24 h day. The juxtaposed cages were mirrored so that the creep area was on the left in one and on the right in the other. Each lactating sow was monitored, i.e., with progeny present in the crate, over two to three successive days between day 6 and

day 23 of the lactation period. For monitoring, Bascom cameras were used; fixed on a metal support 4.5 m above the ground, with its lens pointing downward and directly above each crate to obtain a rear view of the sow with a tilt angle between approximately 45° and 60° (Fig. 1). The cameras were connected to a PC via a wire network for data storage. The videos were manually labelled by three observers, who were trained in using the same ethogram (Table 1).

Dedicated software was implemented in Python language to play the recorded videos and manually report postural changes and the name of the new posture. The function automatically recorded the frame number of the change and the posture was derived automatically between two changes. It was possible to rewind and slow down the video recording to capture changes in posture accurately, annotating when the sow was in a transitory posture (sitting or kneeling) between two main postures (standing and lying). The third observer analysed videos without the dedicated Python software and recorded information on the sow's posture

manually. In this case, rewinding and slowing down the video recording were again used to capture the information accurately. For the purpose of annotation, eight different postures were distinguished: standing, sitting, kneeling, sternal lying, lateral lying on the right or the left side, and with udder exposed (i.e., accessible to the piglets) or not. Only four postures were used for statistical analysis of the signal: standing, sitting, sternal, and lateral. The sternal posture includes sternal and lateral lying on the right or the left side. The lateral posture is used when the sow is lying on its right or left side with the udder visible (see Fig. 1). The animals were monitored during staff working hours, i.e., between 8 AM and 6 PM.

2.2. Image processing

As shown in Figure 1(a), the initial images were cropped to remove the area outside the cage, including piglet creep area, which was not needed to determine sow posture.
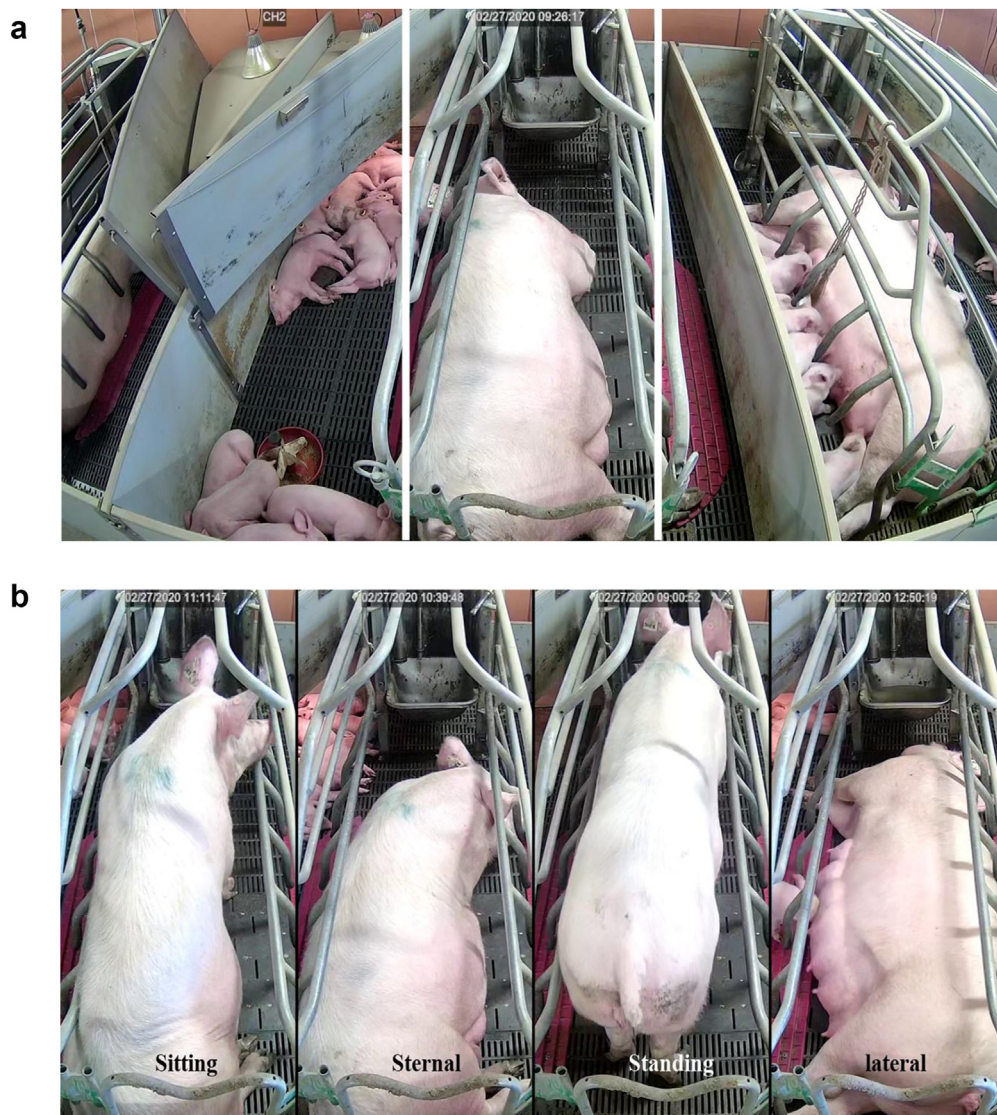
Fig. 1 — Image examples. (a) The initial image captured by the camera. Only the image included between the white lines was considered for the analysis. (b) Example of the four different sow postures.

Table 1 – Description of the ethogram used to label videos.

Standing	Standing up on its four legs
Sitting	Sitting with front legs straight and both front hooves on the ground
Lying sternally	Lying with udder hidden under the belly, including sternally (i.e., with the backbone visible in the centre of the body and the front legs under the sow) or laterally on the right or the left side with udder not visible
Lying laterally	Lying on its right or left side with the udder visible, i.e., accessible to the piglets

The posture prediction model (PPM) has two compartments. First, a feature extraction method (FEM) was used to extract the features from the images. Second, a SVM classification model (SCM) was used to estimate the sow posture, based on the values of the features. Thus, a PPM is composed of one FEM and one SCM. Two types of FEM were considered, Seg-FEM, when the features were extracted using segmentation and CNN-FEM, when the features were extracted using a CNN model. Note that the parameters of the CNNs were not retrained. Segmentation also did not require any training, thus both FEM can be considered as unsupervised. Only the SCM required training in both cases.

2.2.1. Extraction of segmentation based features

The image was segmented using an adaptive thresholding method. The initial RGB image was first converted into an intensity image and was then shrunk by a factor of 2 using the bicubic interpolation method. A threshold, denoted t, was computed for each image, and was defined as the 44% quantile of the non-zero intensity values of the image. In other words, 44% of the non-zero pixels had an intensity value $>t$. Pixels $<t$ were set to 0 and pixels higher than or equal to were set to 1. At this stage, a binary image was created in which only the sow and her piglets were visible. Unfortunately, the fences were also segmented. To select only the sow, the area of all the connected components (i.e. blocks of neighbouring pixels of the same colour) was computed and only the element with the biggest surface area, corresponding to the sow, was kept. Before computing the area, morphological operations (erosion followed by dilatation) were used with a 10-pixel square filter.

Finally, the following features were computed on the segmented image: area, convex area, eccentricity, equivalent diameter, Euler number, extent, filled area, major and minor axis length, orientation, perimeter and solidity. Also considered were the number of pixels in the right, left, bottom, middle and top parts of the image, the number of pixels on the diagonal, as well as the number of connected components. The segmentation approach is shown on Figure 2.

2.2.2. Extraction of CNN based features

CNN is a powerful method for detecting and classifying images. CNNs are a succession of different mathematical operations, such as convolutions, simple non-linear functions (*activation* functions) and dimension reduction (*pooling*). Convolutions are filters applied to the image with a sliding window. The aim is to extract useful information or features from the image. The activation and pooling functions select and summarise the most relevant features. From the initial image, the CNN consists in one or several parallel flows made with these mathematical operations that all converge into a final layer containing all the features extracted from the image. The CNN then uses these features to classify the images. Unlike segmentation, the extracted features are not intuitive and cannot be easily interpreted. Thus, designing and training a CNN from scratch can be very challenging. Fortunately, for image analysis, several CNNs, generally designed by companies such as Google, are available free of charge. These CNNs are trained to recognise common objects, like bikes, cars, or dogs on more than a million images and evaluated on hundreds of thousands of images. Even though they cannot be used directly to recognise a particular type of object or scene, such as sow postures, the features extracted from CNNs are a useful source of information, and today such information is increasingly used to retrain classification methods.

Considered were 11 different CNNs: Xception (Chollet, 2017), InceptionV3 (Szegedy et al., 2016), GoogleNet (Szegedy et al., 2015), MobileNetV2 (Sandler et al., 2018), ResNet50 and ResNet18 (He et al., 2016), DarkNet53 (Redmon, n.d.), DenseNet201 (Huang et al., 2017), AlexNet (Krizhevsky et al., 2012), SqueezeNet (Iandola et al., 2016) and ShuffleNet (Zhang et al., 2018). All the CNN were already pre-trained on the ImageNet dataset (Deng et al., 2009) and were downloaded thought the MATLAB 2020b add-on Explorer. For each CNN, the extracted features were defined as the output value of the last pooling layer. For SqueezeNet for example, the features was the output of the layer named *pool10*. More details on CNN can be found in the literature, such for example in (Wang et al., 2018) or in (Khan et al., 2020).

CNNs were already pre-trained on the ImageNet dataset (Deng et al., 2009) and were downloaded using the MATLAB 2020b add-on explorer. For each CNN, the extracted features were defined as the output value of the last pooling layer. For example, for SqueezeNet, the features were the output of the layer named pool10. More details on CNN can be found for example in (Wang et al., 2018) and (Khan et al., 2020).

Several characteristics of the FEMs are listed in Table 2.

2.3. Posture prediction from features - SVM classification model

Segmentation and CNN both provide a set of features, i.e., a set of variables that are supposed to describe the image and help differentiate between the different sow postures. To predict posture from features, we used an SVM classifier. For the CNN-FEM, we used a multiclass Error-Correcting Output Codes (ECOC) model, which breaks a k class classification problem down into $k(k-1)/2$ binary classification problems. A linear SVM classifier is then fitted for each binary classification problem (Allwein et al., 2000). For the Seg-FEM, we used regular SVM, with an order-3 polynomial kernel, as the ECOC model produced poor results. All the computation in this article was carried out using MATLAB 2020b.

Fig. 2 – Segmentation example. On the left is the initial image together with the segmentation result. What is estimated to be the sow is delimited with a white line. On the right is the result of the adaptive thresholding method. Bars are deleted by keeping the connected components with the highest area.

Table 2 – Characteristics of the feature extraction methods. Computing times are provided for a Precision 3640 Dell Tower, with an intel(R) Core (TM) i9 CPU at 3.70GhZ, a Quadro P2200 GPU, running on Windows 10 Professional. Algorithms were developed in MATLAB 2020b. Depth is the number of convolutional and fully connected layers. Training and Testing times are in seconds for 100 images.

	Xception	InceptionV3	GoogleNet	MobileNetV2	ResNet50	ResNet18
Parameters (10^6)	22.9	23.9	7	3.5	25.6	11.7
Depth	71	48	22	53	50	18
Features	2048	2048	1024	1280	2048	512
Training Time	2.211	0.925	0.354	0.514	0.646	0.294
Testing Time	2.594	1.672	0.66	1.036	1.117	0.691
	DarkNet53	DenseNet201	AlexNet	SqueezeNet	ShuffleNet	Segmentation
Parameters (10^6)	41.6	20	61	1.24	1.4	.
Depth	53	201	8	18	50	.
Features	1024	1920	9216	1000	544	18
Training Time	1.71	1.854	1.717	0.222	0.25	0.14
Testing Time	2.415	3.076	0.524	0.562	0.758	0.01

2.4. Training and test datasets

2.4.1. Initial dataset

Video records of the 12 sows were used, which represented a total of 129 h of manually labelled video. A total of 7.8 h of video were labelled for Sow 1 on day 16 of lactation, 8.3 h for Sow 2 on day 20, 8.2 h for Sow 3 on day 14, 8.6 h for Sow 4 on day 12, 7.8 h for Sow 5 on day 21, 13.4 h for Sow 6 on day 8 and day 9, 17.5 h for Sow 7 on day 5 and day 6, 26.7 h for Sow 8 from day 6 to day 8, 7.2 h for Sow 9 on day 12, 7.7 h for Sow 10 on day 16, 7.8 h for Sow 11 on day 20 and 7.8 h for Sow 12 on day 20. Sows 6 and 7 were labelled for two consecutive days, and Sow 8 was recorded and labelled for three consecutive days. However, problems were encountered during the second day of the labelling for Sow 6, and consequently only labelling done on the first day for this sow was used. For Sow 2 and Sow 12, a radar was installed on top of the crate to estimate sow posture with this technology for the purposes of another study. As a result, these sows' heads were not visible when the head was outside the zone covered by the radar. Sow 1 and Sow 4 had blue marks on their bodies. Because the mounting of the cameras was not standardised, the angle of view differed slightly for each sow and unfortunately for some

sows the cameras were accidently moved slightly during the recording period. A sample image for each sow is provided in supplementary material S2. From the labelled videos, which were recorded at 10 frames s^{-1}, one image was extracted every 10 s. After extraction, the dataset included 41,568 frames. In general, the sternal and lateral postures (resp. 33.9% and 40.6%) were well represented in the dataset, while the standing and sitting postures were not (13.4% and 12.1% respectively). The number of images for each posture and animal is listed in supplementary material S3. This initial dataset was used to construct different training and testing datasets to investigate the impact of the training set on the quality of the posture prediction model.

2.4.2. Baseline scenario: training and testing from the same bank of images

In this baseline scenario, the posture prediction models were trained and tested on a set of images chosen randomly from the initial dataset, without considering the sow number. Two different training sizes, 5% and 30% of the initial dataset, were used. In each case, the training images were chosen randomly and could thus originated from any of the monitored animals, for any posture. The posture prediction models were then

tested on the remaining images not used for training. Only the first day of recording of the animals was used. In this scenario and in the other scenarios, the results obtained were compared when only one posture prediction model was fitted for <u>all</u> animals (group-model) and when one posture prediction model was fitted <u>per</u> animal (individual-model).

2.4.3. Temporal variations

The impact of varying environmental conditions during monitoring was investigate by, for example, small changes in the camera angle. The experimental set up was very similar to that in the baseline scenario in that only the testing dataset was different. In this case, the testing dataset was based on the second monitoring day of sow 7 and on the second and third recording days of sow 8. For the individual posture prediction model, the model that was fitted on the same sow was used. For example, to evaluate the second day of monitoring sow 7 in the case of an individual posture prediction model, we used the model fitted on 5% or 30% of the first day of monitoring sow 7.

2.4.4. Individual variations

The impact of individual variations on the quality of the posture prediction model was also investigated. As mentioned earlier, the environment differed between sows. For example, a radar was installed on top of the crate for some sows, which were consequently not entirely visible in the resulting image. The camera angles generally differed between sows. Lightning conditions also varied greatly, for example when a sow was located near a window, compared to a sow located far from a window. To investigate the possible impact of these variations on the quality of a posture prediction model, we propose separating the sows used for training from the sows used for testing. In other words, the training image dataset and the testing image dataset should not contain images of the same sow.

Firstly, a severe scenario was considered in which the posture prediction models are trained using the images from one sow and tested using the images of all the other 11 sows. Note that in this case, the group-model and individual-model are the same, in that training is performed using images of the same sow. Only the case where 30% of the available data were used for training was considered relevant. Also considered was the opposite scenario, named the soft scenario, where the posture prediction models were trained on all but one sow and tested on the sow not included in the training. In this case, the group and individual posture prediction models differed.

In each scenario, 30% of the available images were used for training. All images available for testing were used. For example, the severe scenario was started by randomly selecting 30% of the images available for sow number 1 and trained all the posture prediction models considered. Then all the images available were used for sow 2, sow 3, etc., up to sow 12, to evaluate the posture prediction models.

In any given case, only the first day of monitoring of each sow was used.

For a given posture $p \in \{Standing, Sitting, Sternal, Lateral\}$, we used the f1-score to evaluate the qualities of the posture prediction model:

$$f1 - score_p = 100 \times \left(2 \times \frac{Sensitivity_p \times Precision_p}{Sensitivity_p + Precision_p} \right)$$

$$Sensitivity_p = \frac{TP_p}{TP_p + FN_p} \tag{1}$$

$$Precision_p = \frac{TP_p}{TP_p + FP_p}$$

where TP_p, TN_p, FP_p and FN_p are the number of true positive, true negative, false positive and false negative detections for posture p. The f1-score is the harmonic mean of precision and sensitivity. Note that a null precision and/or a null sensitivity gives a null f1-score. The sensitivity, also called recall, is the true positive rate for a given posture. For example, a sensitivity of 70% for sternal posture meant that if the animal was in sternal posture, then the model predicted the posture correctly in 70% of the cases. The precision is the positive predicted value. For example, a precision of 70% for sternal posture means that out of all the images predicted as sternal posture, in 70% of them, the sow was truly in sternal posture; but in 30% of the images, the animal was in another posture. To summarise the quality of the method, we calculated the average f1-score as the average of the four postures f1-scores:

$$f1 - score = \frac{\sum_p f1 - score_p}{4}. \tag{2}$$

It should be noted that for all the experimental setups, the training and testing dataset were the same for all the posture prediction models. For the experimental setups where the posture prediction models are trained and tested on several datasets, the average f1-score was calculated and used in the statistical analysis. For example, to study temporal variations with individual model, the f1-score calculated for the second day of monitoring of sow 7, and for the second and third monitoring days of sow 8 were averaged.

3. Results

The results of all the posture prediction models and experimental setups are listed in Table 3. Results are also summarised in Figure 3.

3.1. Baseline scenario: training and testing using images from the same image bank

For the baseline scenario with one model for all animals, the segmentation model produced the lowest f1-score. The score was 89% when 5% of the data were used for training and 93.3% when 30% of the data were used for training. Among all the CNN-PPM, the one based on AlexNet had the highest f1-score in all cases. With 5% of the data used for training, the f1-score was 97.1% and with 30% training, it was 97.7%. It should be noted that increasing the size of the training data set had only a small impact on the model f1-score, compared to the segmentation model. With the MobileNetV2-PPM and ResNet50-PPM, the f1-score did not even change. The GoogleNet-PPM was the most affected by the increased size

Table 3 – Average F-Score for the different methods and setups. For a given experimental setup, the highest F-Score is bold.

Baseline scenario: training and testing from the same bank of images											
Group-PPM: One model for all animals											
5% 93.7	94.2	93.8	94.4	95.3	93.9	94	95.3	**97.1**	94.5	94.4	89
30% 95	95	95.6	94.4	95.3	95.3	94.8	94.5	**97.7**	95.7	94.9	93.3
Individual-PPM: One model per animal											
5% 90.3	89.6	91.9	91.5	91.4	91.4	89.5	92.3	**92.6**	91.2	91.8	86.9
30% 94.3	94.7	95.4	94.4	95.1	95.7	95	95.8	**96.3**	95.2	95.7	92.6
Temporal variations: training on one recording day, testing on others Group-PPM: One model for all animals											
5% 80.8	78.7	77.2	82	82.2	75	**83.4**	78.1	82.4	78.3	80.1	60.7
30% 81.4	77.7	76.9	78.4	79.3	77.4	83.5	76.7	**86.7**	79.2	81.3	64.9
Individual-PPM: One model per animal											
5% 62.3	65	71.1	69	61.2	76.7	65.3	73.1	69.7	68.2	**79.8**	65.7
30% 65.6	60.1	72.2	74.4	57.4	76.9	62	64.1	61.7	74.9	**82.4**	67.2
Individual variations: training and testing on different sows Severe scenario: Training on one sow (30%), test on all other sows											
61	61.8	52.4	62.3	60.1	61.1	60.4	**63.4**	59	55.1	56	38.8
Soft scenario: Training on all animals but one (5% or 30%), test on the non trained animal											
5% 77.5	79.1	77.9	78.7	80.2	75	80.2	82.8	**86.2**	81.2	82.8	59.7
30% 76.8	78.3	78.3	75.1	78.9	76.3	77.6	77	**86**	76.8	76.1	62.2

of the training data set. Its f1-score changed from 93.8% with 5% training to 95.6% with 30% training. Finally, the CNN model based on DenseNet201 was the only one for which the increased size of the training data set reduced the f1-score. With 5% training, the f1-score was 95.3% and 94.5% with 30% training. With the Seg-PPM, the sternal and lateral postures remained the hardest postures to predict. This is surprising given that the image of the sow is supposed to be wider when lying laterally, it should have helped differentiation between these two postures. However, when the animals were in the sternal posture, they may tend more to the right or to the left, which can lead to confusion. Some confusion also appeared when the sow was sitting. When the sow was perfectly positioned on its sternum, or sitting or standing, only the back of the animal was visible and thus may be segmented, resulting in minimum variation in shape between the different postures. There was also some confusion between the sternal and standing postures with the CNN-PPMs, almost certainly for the same reasons. When one posture prediction model was fitted per animal, the CNN-PPM based on AlexNet also produced the best results, with an f1-score of 92.6% with 5% training and of 96.3% with 30% training. Generally, fitting the posture prediction model for only one animal lowered the f1-score, certainly due to the fact that fewer data are available from training. Nonetheless, the f1-score with 30% training was slightly higher with the CNN-PPMs based on ResNet18, DarkNet53, DenseNet201 and SuffleNet.

3.2. Temporal variations

When posture prediction model was fitted for all animals with 5% training, the CNN-PPM based on DarkNet53 produced the best results, with an f1-score of 83.4%. With 30% training, the CNN MODEL based on AlexNet produced the best results, with

an f1-score of 86.7%. On average across all the posture prediction models, the f1-score decreased by 15.9% compared to the baseline scenario with one posture prediction model for all animals and 5% training. With 30% training, it decreased by 16.5%. When one posture prediction model was fitted per animal, the CNN model based on SuffleNet produced the best results, with an f1-score of 79.8% with 5% training and of 82.4% with 30% training. Note that in this experimental setup, with 5% training, the Segmentation model f1-score was 65.7%, i.e., higher than in the CNN models based on Xception (62.3%), InceptionV3 (65%), ResNet50 (61.2%) and DarkNet53 (65.3%). With 30% training, the Segmentation model f1-score was 67.2%. Again, it was higher than that of 6 of the 11 CNN models: Xception (65.6%), InceptionV3 (60.1%), ResNet50 (57.4%), DarkNet53 (62%), DenseNet201 (64.1%) and AlexNet (61.7%). As one can see, there was a difference of nearly 10% with ResNet50. On average over all the posture prediction models tested, the f1-score decreased by 22% compared to the baseline scenario with one posture prediction model per animal and 5% training; with 30% training, it decreased by 26.8%.

3.3. Individual variations

In the severe scenario, the CNN model based on DenseNet53 produced the highest f1-score: 63.4%. With the Segmentation model, the score only reached 38.8%, which is significantly lower than that obtained with the CNN model. Compared to the baseline scenario with 30% training the f1-score was reduced by an average of nearly 37.5% with either the baseline group models or the baseline individual models. In the soft scenario, the CNN model based on AlexNet produced the highest f1-score: 86.2% with 5% training and 86% with 30% training. With most CNN models, increasing the size of the training dataset reduced the f1-score, whereas this was not the case with the Segmentation model. With 5% training, the
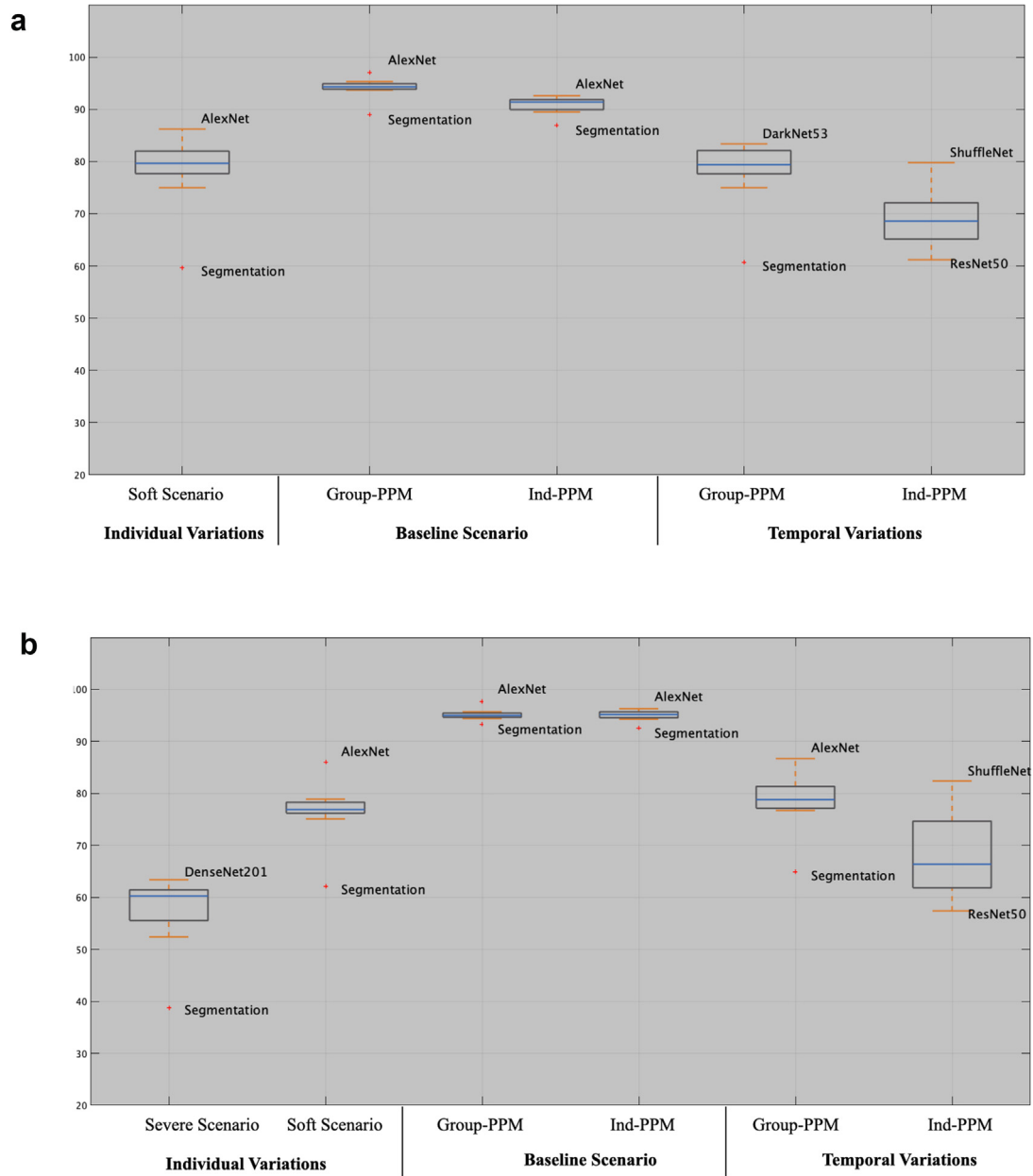
Fig. 3 — Summary of the results. The boxplots give the distribution of the f1-score of all the PPM. There is one boxplot per experimental setup. For each experimental setup, the name of the worst and best posture prediction model is provided. The blue lines indicates the median value and the orange horizontal lines indicate the 25th and 75 percentiles. Figure (a) gives the results when 5% of the data was used for training and 30% for Figure (b). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

f1-score decreased by on average, 16%, compared to the baseline group models, and by 16% with 18.5% training.

4.   Discussion

Different ways of extracting features from images have been investigated. Two type of feature extraction methods (FEM) were used: Seg-FEM, where the features represented geometric characteristics of the segmented image, and CNN-FEM, where the features were calculated as the output of the final pooling layer of the CNN model.

The different f1-scores of the posture prediction models were in line with those obtained in other similar studies. Nasirahmadi et al., (2019) developed a method based on image segmentation and an SVM classifier, similar to the one that we used. In their case, posture was estimated on several pigs kept in a single pen and only the lateral and sternal postures were considered. One group prediction model was trained using 28,035 images and tested on 12,015 images (i.e., 70% training, 30% testing). The f1-score of the prediction model was estimated to be 94.2%. This is comparable to the segmentation model in our baseline scenario with 30% training, in which the f1-score was 93.3%. Leonard et al., (2019) developed an

unsupervised method to distinguish between sitting, standing, kneeling, and lying, for sows in farrowing stalls, using top view depth images acquired with a Microsoft Kinect V2 system. The f1-score was estimated to be 79.2% on a dataset of 24 individuals, for each of whom 445 images were chosen randomly. But the f1-score is particularly reduced by the kneeling posture. Similar to our study, Kasani et al. (2021) evaluated eight different CNN models, with 7556 training images and 1776 images used for testing (80% training 20% testing). They differentiated four postures: lying left, lying right, sitting and standing. All the CNNs tested provided a f1-score higher than 99%. The best tested CNNs were DenseNet121 and DenseNet201. The difference in the f1-score compared to our analysis could be explained by the difference in the postures considered. The sternal posture, which was regularly confused with the lateral postures, was accounted for. In the work by Kasani et al. (2021), the sternal posture was not accounted for. More generally, it is difficult to compare different works rigorously, because postures were not necessarily defined in exactly the same way and also because the environment, the sensors, and the position of the sensors may differ. As shown in our work, the choice of the training and test sets can also greatly influence the results.

In the present study, annotation relied on the continuous analysis of video recordings for several hours per sow. After the image labelling process, the labelled images were double-checked to correct any errors. Observed discrepancies occurred between the behaviour and its annotation on a minor but substantial proportion of data. To some extent, this was due to a delay between human observation and pressing the computer key, and the limited use of backtracking. Random sampling of video images would probably have been more efficient.

To improve the quality of prediction, future work could focus on aggregating information from different sensors or methods. Interestingly, the CNN can also return a probability for different postures. Therefore, several cameras, with different angles and axes, can be used and analysed by different CNNs. The estimated probabilities could then be combined to estimate the postures. In addition, depending on the frequency of image acquisition, temporal information can be added to facilitate the detection of postures; for example, using the transition probability between postures or to smooth the postural signal. Finally, features from several FEMs can be aggregated to construct a meta-posture prediction model.

This work raises two important questions. When does the posture prediction model need to be retrained when a new batch of sows is monitored? Should labelling images wait until the monitoring period ends so as to include images from all the monitoring days in the training dataset? To answer these questions correctly, a larger dataset is needed.

5. Conclusions

Several conclusions can be drawn from this work. Firstly, in general it is not necessary to train one model per animal. When one posture prediction model was used for all sows, the training dataset was larger in size and more diverse, which certainly explained the better performance of this approach. In practice, it is more convenient to build only one training dataset of labelled images. Secondly, the CNN models generally produced a higher quality of prediction than the Segmentation model. However, the Segmentation model may be more robust to temporal variations than some CNN models when the model is trained individually. When the training and testing images were selected from the same pool of images, the difference between Segmentation model and CNN models was relatively small. The Segmentation model f1-score was 1% lower than the worst CNN model and 2% lower than the best CNN model. The Segmentation model is thus a good option if the CNN models are too complex to implement. Thirdly, an average difference of 9.7% was observed between the f1-score of the best and worst CNN-PPM. It is thus appropriate to test several CNNs for a given application. Fourthly, it is important to train the posture prediction model on a dataset that includes images similar to those for which the posture is to be predicted. For example, we showed that the f1-score decreased on average by 18.5% when the posture prediction model was used on a sow that was not present in the training set. For temporal variations, due to, for example, a change in sow behaviour, in lightning conditions, or in the field of view of the camera, it was seen that the f1-score decreased by an average of 16%.

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.

Acknowledgement

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.biosystemseng.2021.09.014.

REFERENCES

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, J., Kaiser, L., Kudlur, M., et al. (2015). TensorFlow: Large-scale machine learning on heterogeneous systems. http://tensorflow.org/.

Allwein, E. L., Schapire, R. E., & Singer, Y. (2000). Reducing multiclass to binary: A unifying approach for margin classifiers. Journal of Machine Learning Research, 1, 113–141.

Andersen, I. L., Berg, S., & Bøe, K. E. (2005). Crushing of piglets by the mother sow (Sus scrofa)—purely accidental or a poor mother? Applied Animal Behaviour Science, 93(3–4), 229–243. https://www.sciencedirect.com/science/article/pii/S0168159104003065.

Aydin, A., Cangar, O., Eren Ozcan, S., Bahr, C., & Berckmans, D. (2010). Application of a fully automatic analysis tool to assess the activity of broiler chickens with different gait scores. *Computers and Electronics in Agriculture, 73*(2), 194–199.

Baxter, E. M., Andersen, I. L., & Edwards, S. A. (2018). Sow welfare in the farrowing crate and alternatives. In *Advances in pig welfare* (pp. 27–72). Elsevier.

Bezen, R., Edan, Y., & Halachmi, I. (2020). Computer vision system for measuring individual cow feed intake using RGB-D camera and deep learning algorithms. *Computers and Electronics in Agriculture, 172*, 105345. https://www.sciencedirect.com/science/article/pii/S0168169919313249.

Bonneau, M., Vayssade, J. A., Troupe, W., & Arquet, R. (2020). Outdoor animal tracking combining neural network and time-lapse cameras. *Computers and Electronics in Agriculture, 168*, 105150. https://www.sciencedirect.com/science/article/pii/S0168169919322562.

Brown, D. D., Kays, R., Wikelski, M., Wilson, R., & Peter Klimley, A. (2013). Observing the unwatchable through acceleration logging of animal behavior. *Animal Biotelemetry, 1*(1), 20. https://link.springer.com/article/10.1186/2050-3385-1-20.

Canario, L., Bidanel, J. P., & Rydhmer, L. (2014). Genetic trends in maternal and neonatal behaviors and their association with perinatal survival in French large white swine. *Frontiers in Genetics, 5*, 410. https://www.frontiersin.org/articles/10.3389/fgene.2014.00410/full.

Canario, L., Labrune, Y., Bompa, J. F., Billon, Y., Ravon, L., Reignier, S., … Ricard, E. (2019). Development and validation of an embedded tool to measure postural activity of lactating sows. *Journées de la Recherche Porcine en France, 50*, 329–330. https://hal.inrae.fr/hal-02737266.

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. *Behavioural Processes*, 104262. https://www.sciencedirect.com/science/article/pii/S0376635720304551.

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. *Computers and Electronics in Agriculture, 169*, 105166. https://www.sciencedirect.com/science/article/pii/S0168169919319556.

Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition (Seattle)* (pp. 1251–1258).

Commision European. (2021). Communication from the commission on the European citizens' intiative (ECI) end the cage edge. https://ec.europa.eu/transparency/documents-register/detail?ref=C(2021)4747&lang=en.

CWF. (2021). Compassion in world farming - scientific briefing on caged farming overview of scientific research on caged farming of laying hens, sows, rabbits, ducks, geese, calves and quail. https://www.europarl.europa.eu/cmsdata/231963/Scientific%20briefing%20on%20caged%20farming,%20February%202021.pdf.

Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition* (pp. 248–255). http://www.image-net.org.

Dwyer, C. M. (2014). Maternal behaviour and lamb survival: From neuroendocrinology to practical application. *Animal: An International Journal of Animal Bioscience, 8*(1), 102.

Escalante, H. J., Rodriguez, S. V., Cordero, J., Kristensen, A. R., & Cornou, C. (2013). Sow-activity classification from acceleration patterns: A machine learning approach. *Computers and Electronics in Agriculture, 93*, 17–26.

Gan, H., Ou, M., Zhao, F., Xu, C., Li, S., Chen, C., & Xue, Y. (2021). Automated piglet tracking using a single convolutional neural network. *Biosystems Engineering, 205*, 48–63. https://www.sciencedirect.com/science/article/pii/S1537511021000465.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *2016 proceedings of the IEEE conference on computer vision and pattern recognition (Seattle)*.

Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In *2017 proceedings of the IEEE conference on computer vision and pattern recognition (Seattle)*.

Iandola, F. N., Han, S., Moskewicz, M. W., Ashraf, K., Dally, W. J., & Keutzer, K. (2016). *SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and<0.5 Mb model size.* arXiv Preprint arXiv:1602.07360.

Kasani, P. H., Oh, S. M., Choi, Y. H., Ha, S. H., Jun, H., Park, K. H., … Kim, J. S. (2021). A computer vision-based approach for behavior recognition of gestating sows fed different fiber levels during high ambient temperature. *Journal of Animal Science and Technology, 63*(2), 367. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8071751/.

Kashiha, M., Bahr, C., Haredasht, S. A., Ott, S., Moons, C. P. H., Niewold, T. A., Ödberg, F. O., & Berckmans, D. (2013a). The automatic monitoring of pigs water use by cameras. *Computers and Electronics in Agriculture, 90*, 164–169. https://doi.org/10.1016/j.compag.2012.09.015

Kashiha, M., Pluk, A., Bahr, C., Vranken, E., & Berckmans, D. (2013b). Development of an early warning system for a broiler house using computer vision. *Biosystems Engineering, 116*(1), 36–45.

Khan, A., Sohail, A., Zahoora, U., & Qureshi, A. S. (2020). A survey of the recent architectures of deep convolutional neural networks. *Artificial Intelligence Review, 53*(8), 5455–5516.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems, 25*, 1097–1105.

Leonard, S. M., Xin, H., Brown-Brandl, T. M., & Ramirez, B. C. (2019). Development and application of an image acquisition system for characterizing sow behaviors in farrowing stalls. *Computers and Electronics in Agriculture, 163*, 104866. https://www.sciencedirect.com/science/article/pii/S0168169919305666.

Li, W., Ji, Z., Wang, L., Sun, C., & Yang, X. (2017). Automatic individual identification of holstein dairy cows using tailhead images. *Computers and Electronics in Agriculture, 142*, 622–631. https://www.sciencedirect.com/science/article/pii/S0168169917300649.

Marsot, M., Mei, J., Shan, X., Ye, L., Feng, P., Yan, X., Li, C., & Zhao, Y. (2020). An adaptive pig face recognition approach using convolutional neural networks. *Computers and Electronics in Agriculture, 173*, 105386.

Matheson, S. M., Thompson, R., Ploetz, T., Kyriazakis, I., Walling, G. A., & Edwards, S. A. (2017). *Relationships between sow conformation, accelerometer data and crushing events in commercial piglet production. In 51st annual meeting of the international society for applied ethology.* UK: Newcastle University. https://www.researchgate.net/profile/Stephanie_Matheson/publication/320331371_Relationships_between_sow_conformation_accelerometer_data_and_crushing_events_in_commercial_piglet_production/links/59de2ddda6fdcca0d32034b2/Relationships-between-sow-conformation-accelerometer-data-and-crushing-events-in-commercial-piglet-production.pdf.

Moses, & Olafenwa, J. (2018). *ImageAI, an open source Python library built to empower developers to build applications and systems with self-contained computer vision capabilities.* https://github.com/OlafenwaMoses/ImageAI.

Nalon, E., Maes, D., Van Dongen, S., van Riet, M. M. J., Janssens, G. P. J., Millet, S., & Tuyttens, F. A. M. (2014).

Comparison of the inter-and intra-observer repeatability of three gait-scoring scales for sows. *Animal: An International Journal of Animal Bioscience, 8*(4), 650. https://pdfs.semanticscholar.org/7a99/992b32033d79cdc4b9790717d599614f786c.pdf.

Nasirahmadi, A., Edwards, S. A., & Sturm, B. (2017). Implementation of machine vision for detecting behaviour of cattle and pigs. *Livestock Science, 202,* 25–38. https://doi.org/10.1016/j.livsci.2017.05.014

Nasirahmadi, A., Sturm, B., Olsson, A. C., Jeppsson, K. H., Müller, S., Edwards, S., & Hensel, O. (2019). Automatic scoring of lateral and sternal lying posture in grouped pigs using image processing and support vector machine. *Computers and Electronics in Agriculture, 156,* 475–481. https://www.sciencedirect.com/science/article/pii/S0168169918313516.

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://www.mdpi.com/2076-2615/9/8/538.

Oczak, M., Maschat, K., Berckmans, D., Vranken, E., & Baumgartner, J. (2016a). Automatic estimation of number of piglets in a pen during farrowing, using image analysis. *Biosystems Engineering, 151,* 81–89. https://www.sciencedirect.com/science/article/pii/S1537511016303464.

Oczak, M., Maschat, K., Berckmans, D., Vranken, E., & Baumgartner, J. (2016b). Can an automated labelling method based on accelerometer data replace a human labeller? − postural profile of farrowing sows. *Computers and Electronics in Agriculture, 127,* 168–175. https://doi.org/10.1016/j.compag.2016.06.013

Pedersen, L. J., Jørgensen, E., Heiskanen, T., & Damm, B. I. (2006). Early piglet mortality in loose-housed sows related to sow and piglet behaviour and to the progress of parturition. *Applied Animal Behaviour Science, 96*(3−4), 215–232. https://www.sciencedirect.com/science/article/pii/S0168159105001504.

Redmon, J.. Darknet: Open source neural networks in c. n.d https://pjreddie.com/darknet.

Ringgenberg, N., Bergeron, R., & Devillers, N. (2010). Validation of accelerometers to automatically record sow postures and stepping behaviour. *Applied Animal Behaviour Science, 128*(1−4), 37–44. https://www.sciencedirect.com/science/article/pii/S0168159110002674.

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L.-C. (2018). MobileNetV2: Inverted residuals and linear bottlenecks. In *2018 proceedings of the IEEE conference on computer vision and pattern recognition (Seattle)*.

Shi, C., Chen, Teng, G., & Li, Z. (2016). An approach of pig weight estimation using binocular stereo system based on labview. *Computers and Electronics in Agriculture, 129,* 37–43. https://www.sciencedirect.com/science/article/pii/S0168169916306329.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A. (2015). Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition (Seattle)*.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In *2016 proceedings of the IEEE conference on computer vision and pattern recognition (Seattle)*.

Thompson, R. J., Matthews, S., Plötz, T., & Kyriazakis, I. (2019). Freedom to lie: How farrowing environment affects sow lying behaviour assessment using inertial sensors. *Computers and Electronics in Agriculture, 157,* 549–557. https://www.sciencedirect.com/science/article/pii/S0168169917312681.

Turner, S. P., & Lawrence, A. B. (2007). Relationship between maternal defensive aggression, fear of handling and other maternal care traits in beef cows. *Livestock Science, 106*(2−3), 182–188. https://www.sciencedirect.com/science/article/pii/S1871141306002861.

Viazzi, S., Ismayilova, G., Oczak, M., Sonoda, L. T., Fels, M., Guarino, M., Vranken, E., Hartung, J., Bahr, C., & Berckmans, D. (2014). Image feature extraction for classification of aggressive interactions among pigs. *Computers and Electronics in Agriculture, 104,* 57–62.

Vizcarra, J. A., & Wettemann, R. P. (1996). Reproducibility, repeatability, and degree of expertise required to assess body condition score in beef cattle. *The Professional Animal Scientists, 12*(1), 28–31. https://www.sciencedirect.com/science/article/abs/pii/S1080744615324785.

Wallenbeck, A., Rydhmer, L., & Thodberg, K. (2008). Maternal behaviour and performance in first-parity outdoor sows. *Livestock Science, 116*(1−3), 216–222. https://www.sciencedirect.com/science/article/pii/S1871141307005136.

Yang, A., Huang, H., Zhu, X., Yang, X., Chen, P., Li, S., & Xue, Y. (2018). Automatic recognition of sow nursing behaviour using deep learning-based segmentation and spatial and temporal features. *Biosystems Engineering, 175,* 133–145. https://www.sciencedirect.com/science/article/pii/S1537511018304641.

Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Liu, T., … Chen, T. (2018). Recent advances in convolutional neural networks. *Pattern Recognition, 77,* 354–377. https://www.sciencedirect.com/science/article/pii/S0031320317304120.

Yang, A., Huang, H., Yang, X., Li, S., Chen, C., Gan, H., & Xue, Y. (2019). Automated video analysis of sow nursing behavior based on fully convolutional network and oriented optical flow. *Computers and Electronics in Agriculture, 167,* 105048. https://www.sciencedirect.com/science/article/pii/S0168169918313942.

Zhang, X., Zhou, X., Lin, M., & Sun, J. (2018). ShuffleNet: An extremely efficient convolutional neural network for mobile devices. In *2018 proceedings of the IEEE conference on computer vision and pattern recognition (Seattle)* (pp. 6848–6856).

Zhang, Y., Cai, J., Xiao, D., Li, Z., & Xiong, B. (2019). Real-time sow behavior detection based on deep learning. *Computers and Electronics in Agriculture, 163,* 104884. https://www.sciencedirect.com/science/article/pii/S0168169919300080.