A computer vision approach for recognition of the engagement of pigs with different enrichment objects

Chen Chen[a,b], Weixing Zhu[a,*], Maciej Oczak[c,d], Kristina Maschat[c,e], Johannes Baumgartner[d], Mona Lilian Vestbjerg Larsen[b], Tomas Norton[b,*]

[a] School of Electrical and Information Engineering, Jiangsu University, Zhenjiang 212013, Jiangsu, China
[b] Division of Measure, Model & Manage Bioresponses (M3-Biores), KU Leuven, Kasteelpark Arenberg 30, 3001 Leuven, Belgium
[c] Precision Livestock Farming Hub (PLF-Hub), University of Veterinary Medicine Vienna, Austria
[d] Institute of Animal Welfare Science, University of Veterinary Medicine Vienna, Austria
[e] FFoQSI GmbH, Technopark 1C, A-3430 Tulln, Austria

A B S T R A C T

As providing objects that pigs prefer can reduce the occurrence of tail-biting and aggression and consequently improve animal welfare, automatic recognition of pigs' engagement with different objects can have practical value. Therefore, aim of this study was to develop a computer vision based approach that utilised a recurrent neural network-based deep learning algorithm to recognise pig enrichment engagement (EE) behaviours and preliminarily determine the preference to objects. Two pig pens were studied. 1 day of video was recorded in pen 1, which generated 2400 1 s EE and 2400 1 s non-EE episodes. 80% of these data was randomly selected as training set and the remaining 20% as validation set. Moreover, 4 days of video were recorded and used as the test set in pen 2. Firstly, the HSV (Hue, Saturation, Value) colour space-based tracking algorithm was developed to locate object region of interest. Secondly, the convolutional neural network (CNN) architecture InceptionV3 was used to extract spatial features from each frame. These features were input into the long short-term memory (LSTM) framework to extract spatial-temporal features from each episode. Through the fully connected layer, the prediction function *Softmax* was finally used to classify these episodes as EE or non-EE behaviour. In the validation set, the proposed algorithm could recognise EE with blue ball, golden ball and wooden beam with an accuracy of 95.2%, 95.4% and 97.3%, respectively. By shortening the radius of the region of interest into a half of the average length of pig body, the corresponding accuracy could be improved into 96.9%, 97.1% and 97.9%, respectively. In the test set, the proposed algorithm could recognise EE with each of these 3 objects with an accuracy of 96.5%, 96.8% and 97.6%, respectively. The proportion of EE with each of these 3 objects was 75.8%, 6.0% and 18.2%, respectively. These results indicate that the proposed method can be used to recognise EE behaviours of pigs, and halving the radius of the region of interest can improve the recognition accuracy of EE behaviours. Moreover, the preference of pigs to objects based on EE duration were preliminarily determined as blue ball > wooden beam > golden ball. The obtained duration of EE behaviours can help farmers to evaluate the enrichment used and thereby to increase the health and welfare of the pigs in their care. Furthermore, the proposed algorithm has reference value for the classification of the behaviours with similar motion patterns.

1. Introduction

The provision of proper environmental enrichment may reduce the occurrence of tail-biting and aggression in group housed pigs (Lahrmann et al., 2018). For instance, a wooden beam (Larsen et al., 2019) and a chewable rubber ball with protrusions (Telkänranta et al., 2014) have been shown to reduce the occurrence of tail-biting, and a smooth surfaced solid ball can reduce the risk of aggression (Fu et al., 2018). Recognising the engagement of pigs with enrichment and quantifying the time that the pigs spend with the enrichment can bring about two advantages: (1) the animals' preferred objects can be determined which can improve the longer-term benefits of enrichment (Turner et al., 2006) and (2) the enrichment engagement time can serve as an quantitative indicator of positive welfare for pig production (Brown et al., 2018). It is widely known that not all objects have a long-lasting positive effect on animal welfare. Understanding the

engagement of pigs with different objects is necessary to more accurately specify the type of enrichment object and also help a farmer in deciding when to change objects when the animals become bored. Therefore, recognition and quantification of the pigs' engagement with different objects can have value to both research and farming practice.

Presently, evaluation of pigs' engagement with enrichment has mainly been performed through human observation, which is time-consuming, laborious and hard to perform objectively and consistently both within and between observers. Addressing this challenge requires the ability to capture both spatial and temporal information on pig behavioural within the region of the enrichment object. Top view cameras have been demonstrated in the literature to be useful in capturing behavioural information on group housed pigs. However, computer vision technology has not been used to automatically recognise this behaviour. Automatic monitoring of enrichment engagement (EE) through computer vision technology has the advantage of being non-intrusive, less-subjective and uninterrupted and also has the potential to recognise simultaneously occurring behaviours along with those performed towards the enrichment materials provided to the pigs (Nasirahmadi et al., 2017).

Recently, deep learning-based computer vision approaches, especially through convolutional neural networks (CNN), has been widely used for the studies of pig behaviours. Initially, Zheng et al. (2018) recognised 5 postures of a lactating sow (i.e. standing, sitting, sternal recumbency, ventral recumbency and lateral recumbency) by using Faster R-CNN (that shares the convolutional features in a Region Proposal Network (RPN) and a Fast R-CNN (Girshick, 2015)) and obtained sows accurate location. Subsequently, Zhu et al. (2020) proposed an end-to-end refined two-stream Red-green-blue Depth (RGB-D) Faster R-CNN algorithm by fusing RGB-D image features in the feature extraction stage to recognise the above 5 sow postures. Yang et al. (2020) extracted the spatial temporal features mainly by using fully convolutional networks (FCNs) and optical flow analysis and classified these features by using hierarchical classifier to recognise sow drinking, feeding, nursing, moving, medium active and inactive behaviours. Yang et al. (2018) used Faster R-CNN to locate and identify individual pigs and extracted feeding area occupation rate to recognise feeding behaviour. Furthermore, Zhang et al. (2019) proposed a Sow Behaviour Detection Algorithm based on Deep Learning (SBDA-DL) to recognise drinking, urination and mounting of sows. In the above pig behaviour studies based on CNN, the CNN architecture was used to extract spatial features and train individual image frames.

As behaviour performed towards the enrichment mainly manifests as a continuous interaction process between pigs and objects, a computer vision algorithm that considers the spatial-temporal motion patterns in videos is necessary to recognise EE of pigs. Recent studies in the computer vision domain have shown that combining the convolutional neural network (CNN) with a long short-term memory (LSTM) framework can offer a powerful framework for extracting spatial-temporal features (Donahue et al., 2015; Srivastava et al., 2015). Long short-term memory (LSTM) is a commonly used Recurrent Neural Network (RNN) (Hochreiter and Schmidhuber, 1997) that has been widely used for gesture recognition (Tsironi et al., 2017), online handwriting recognition (Nguyen et al., 2018) and text report classification (Banerjee et al., 2019). However, the use of LSTM for automated pig behaviour monitoring studies is still limited. Chen et al. (2020) was one of the first studies to extract the spatial-temporal features by combining the CNN (Visual Geometry Group 16 (VGG16) architecture) and LSTM in order to recognise aggressive video episodes of pigs. In that study, the motion velocity and interaction pattern of aggressive behaviours between adjacent frames changed much faster than that of non-aggressive behaviours, and thus non-aggressive pigs were considered as the background and the entire image in the video was directly used as data for training the model.

There are significant differences between pigs' behaviour towards enrichment and pigs' aggressive behaviour. However, the velocity and interaction pattern of EE behaviour may be similar to that of other non-EE behaviours. Therefore, in developing an algorithm for automated recognition of EE, using the entire image as data to train the model may lead to inaccurate results. As a result, in this paper we develop a HSV (Hue, Saturation, Value) colour space-based tracking algorithm (Gonzalez and Woods, 2007) of pigs performing behaviour towards the enrichment object to further remove the region that is unrelated to EE in an image in the conditions of dim illumination, crowded pigs and dirty objects. On the other hand, in order to ensure the accuracy of recognising EE, this paper applies the InceptionV3 network (Szegedy et al., 2016), which is a CNN architecture with greater depth and width than VGG16.

Hence, the aim of this study is to combine InceptionV3 and LSTM to automatically recognise episodes of EE in pigs. Furthermore, this study aims to preliminarily determine pigs preference to 3 different objects by applying the developed algorithm to calculate the duration of engagement with each object.

2. Materials and methods

2.1. Experimental setup

2.1.1. Video acquisition

The videos were collected at an experimental pig farm of University of Veterinary Medicine Vienna (VetFarmMedau, Pottenstein, Lower Austria, Austria). The pigs were crossbred from Landrace × Large White dams and Piétrain boars. Each pen was 3.50 m × 5.48 m, and it contained an automated Schauer feeding station and a nipple drinker. The fattening pigs used in this study were at the age of 10 weeks (30 kg) and stayed until slaughter (120 kg). An IP camera (GV-BX 1300KV, Geovision Inc., Taipei, Taiwan) locked in protective housing (HEB32K1, Videotec, Schio, Italy) was placed above the pen at the height of 5 m relative to the ground. This camera was used to record RGB videos with a resolution of 1280 × 720 pixels and a frame rate of 30 fps.

Each pen of pigs was provided with a chewable enrichment material (i.e. blue rubber ball with protrusions), an enrichment toy (i.e. golden solid ball with smooth surface) and a natural enrichment material (i.e. wooden beam) hanging on a chain from the side inventory of the pen, according to the type and amount complying with the Austrian animal welfare law.

In order verify whether a small number of data (1 day) can be used to train out the effective model to test more data (4 day), data from the first day after pigs were moved from weaner compartment to fattening compartment in pig pen 1 was used as the training and validation data. Data from the first 4 days after pigs were moved from weaner compartment to fattening compartment in pig pen 2 was used as the test data (Fig. 1). These days were chosen as pigs show the most intense engagement with the enrichment objects on the first day after providing these objects and as these objects were provided to the pens on the day of fattening (Larsen et al., 2019). As pigs mostly rest at night, only 8 h of video from 09:00 to 17:00 were used on each day.

The computer processor was Intel(R) Core(TM) i7-8700 K CPU @ 3.70 GHz with 24 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. The software used for developing the algorithms was Python 3.7.3. CNN and LSTM were implemented on the frameworks of Tensorflow 1.13.1 and Keras 2.2.4, respectively.

2.1.2. Enrichment engagement tracking

In order to reduce the interference of non-enrichment engaged pigs in the EE recognition, a HSV colour space-based tracking algorithm was developed to improve the datasets by removing the pigs outside the EE region of interest. As EE behaviour manifests as a continuous interaction process between piglets and objects, the tracking of engaged pigs in this study was converted into a continuous tracking problem of a

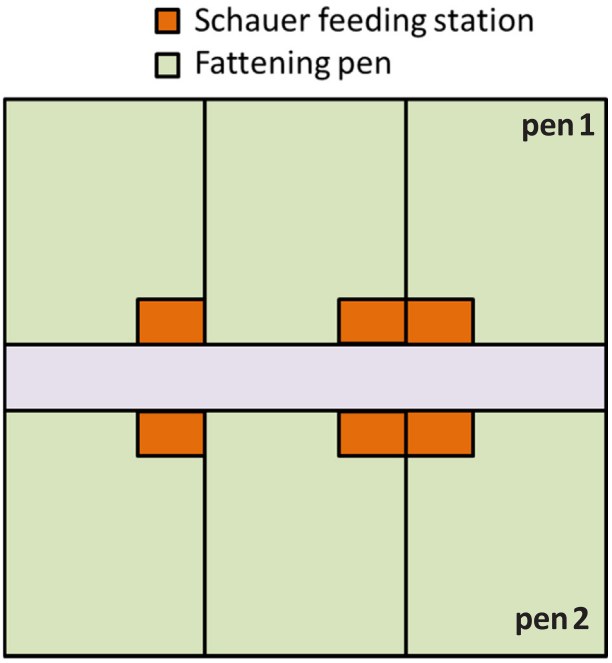

Fig. 1. Layout of the fattening compartment, where pen 1 was located on the first right of the top row and pen 2 on the first right of the bottom row.

circular region around the object. In order to attract the pigs' attention, objects with different colours, shapes and materials were placed in the experimental pens. As the objects have different colours from the pigs and the background, this paper attempts to use the colour characteristic to track the objects. However, the colour of enrichment objects is affected by crowded pigs, dim illumination at the corner of the pen, and dirty object surfaces. Therefore, it is difficult to detect objects directly by using the RGB components of object colours. As the HSV colour space-based object detection method is not sensitive to illumination changes (Gonzalez and Woods, 2007), this paper firstly converted RGB space to HSV space and then used HSV components to locate objects. The specific steps are as follows:

1. Firstly, EE with blue ball was tracked (Fig. 2(I)). Compared to the traditional gray scale-based histogram equalization, the histogram equalization based on the RGB channels of original images (Fig. 2(Ia)) was used to enhance the quality of the images (Fig. 2(Ib)).
2. A series of functions in MATLAB (R2018b, The MathWorks Inc., MA) were used to develop this tracking algorithm. The *rgb2hsv* function was used to transform the RGB space into HSV space (Fig. 2(Ic)). As each colour corresponds to a special range of H, S and V components (Gonzalez and Woods, 2007), the range of the H, S and V components corresponding to blue was set to (0.540, 0.689), (0.169, 1) and (0.180, 1). The *hsv2rgb* function was used to display the blue regions in the HSV space as RGB images (Fig. 2(Id)). From this result, it can be seen that the pigs and the most of the background were removed, and there is an obvious colour difference between the remaining background and the blue ball. In order to further remove background, manual multipoint sampling was performed on blue ball (Fig. 2(Ie)) and set the deviation of the R, G and B values of these standard sample points to 10 pixels to get the extraction result of the blue ball (Fig. 2(If)). In this study, 3000 frames for blue ball being completely visible and in different positions of the pen were performed with manual sampling. The method of collecting sample points is to collect a total of 13 standard sample points at an equal interval of 11 pixels on the rays starting from the centroid at each protrusion.
3. In order to connect these extracted blue points to approximately restore the shape of the ball, the *imdilate* function was used to dilate these points (Fig. 2(Ig)).
4. In order to remove the noise in this result by calculating the area of each connected domain, the *regionprops* and *ismember* functions were used. The largest connected domain was defined as the ball, and other connected domains (i.e. noise) were removed.
5. The centroid of the restored ball was set as the circle center and the average pig length (220pixels) as the radius. This circular region was then used as the region of interest of EE with the blue ball (Fig. 2(Ih)).
6. The method of tracking the region of interest for engagement with the golden ball is the same as steps 1–5 (Fig. 2(II)). In this study, 3000 frames for golden ball being completely visible and in different positions of the pen were performed with manual sampling. The range of the H, S and V components of golden ball was set to (0.080, 0.189), (0.169, 1), and (0.180, 1). The method of collecting sample points is to collect a total of 17 standard sample points at an equal interval of 13 pixels on the rays starting from the centroid of the ball and in the directions of 0°, 45°, 90°, 135°, 180°, 225°, 270°, and 315°.
7. As the wooden beam was hanged on a chain from the pen wall (Fig. 2(III)), the region of interest of EE with the wooden beam was defined as a combination region of a rectangle and 2 quarter circles (Fig. 2(IIIc)). Where r is the average pig length.

2.1.3. Datasets

In the 1 day data of pen1, 7310 episodes of EE with blue ball, 692 episodes of EE with golden ball and 1900 episodes of EE with wooden beam were recorded. It can be seen that the data was unbalanced. The same EE behaviours have great similarity, which is closely related to the number of pigs engaged with objects. As a result, the training and validation sets were built as follows: (1) 400 1 s episodes were selected from the EE with blue ball between 1 and 5 pigs, and 200 1 s episodes were randomly generated from the remaining 6910 episodes. (2) 400 1 s episodes were selected from the EE with golden ball between 1 and 4 pigs, and 200 1 s episodes were randomly generated from the remaining 292 episodes. (3) 400 1 s episodes were selected from the EE with wooden beam between 1 and 3 pigs, and 200 1 s episodes were randomly generated from the remaining 1500 episodes. (4) 600 1 s non-EE episodes for each of the blue ball, golden ball and wooden beam were randomly generated from the non-EE with more than 1 moving pig in the ROI.

In order to verify whether shortening the EE region of interest can be used to improve the performance of the proposed algorithm, two experiments were executed. In experiment 1, in order to keep the balance between the data, 600 1 s EE and 600 1 s non-EE episodes for each of the blue ball, golden ball and wooden beam were labelled on video collected in pen1 (Fig. 3(a)). Video data was augmented by horizontal, vertical and diagonal mirroring with the *imwarp* function in MATLAB and the transformation matrixes of horizontal, vertical and diagonal mirroring (i.e. [−1 0 0; 0 1 0; Width 0 1], [1 0 0; 0–1 0; 0 Height 1] and [−1 0 0; 0–1 0; Width Height 1]). This data augmentation process can be considered to represent the same EE behaviour occurring in 4 different positions and thus different CNN features can be extracted. Among them, 80% of data was randomly selected as training set and the remaining 20% of data as validation set. In the data of pen 2, 115,200 1 s episodes were labelled (32 h, 8 h across 4 days). Out of them, 27,330 1 s episodes were labelled as EE with the blue ball, 2151 1 s episodes were labelled as EE with the golden ball, and 6461 1 s episodes were labelled as EE with the wooden beam. The remaining 1 s episodes were labelled as non-EE with each of these 3 objects. All of these data were used as test set.

Literature has shown that CNN can be used to process the problem of touching pigs. For instance, Tian et al. (2019) modified a counting CNN model based on the architecture ResNeXt to count the number of pigs under conditions of partial occlusion, overlapping and different

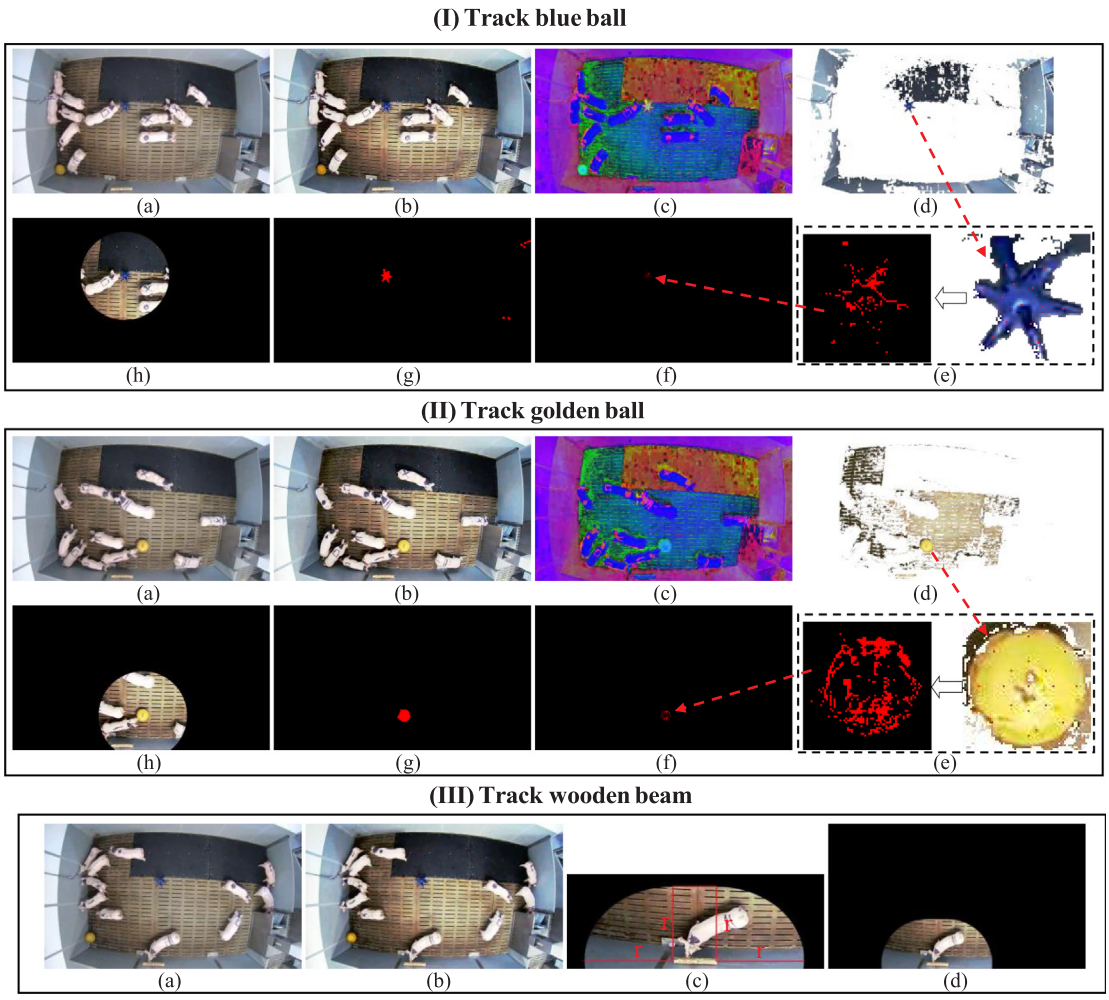

Fig. 2. Tracking process of the region of interest (ROI) of the blue ball, golden ball and wooden beam.

perspectives. However, EE behaviour is mainly concentrated on the object and the front half of the pig, and thus shortening object region of interest (ROI) can further reduce the influence of pig touching on recognition results. As a result, the ROI radius was shortened into a half of average pig length to perform the experiment 2 in this paper (Fig. 3(b)). Among them, the allocation method of training set, validation set and test set is the same as that in Experiment 1.

2.1.4. Labelling

Enrichment engagement behaviour towards the three objects provided to the pig pens is described in Table 1. Furthermore, non-EE behaviours include lying, walking, drinking, feeding, running, chasing and aggression.

Fig. 4 further illustrates the EE frames corresponding to blue ball, golden ball and wooden beam in the labelled data.

2.2. Algorithm

The interaction between enrichment engaged pigs and objects is continuous and demands contact while the interaction between non-enrichment engaged pigs and objects is instantaneous and may do not involve contact. Therefore, LSTM that can extract spatial-temporal features was used to distinguish the difference of this interaction pattern between the engaged and non-engaged behaviour in this study. Since the CNN architecture Inception considers both the depth and width of the network to improve classification performance, the Inception network was input into LSTM to classify EE and non-EE behaviours. Although Inception networks include InceptionV1, InceptionV2, InceptionV3, InceptionV4 and Inception-ResNet-V2, Keras currently only supports transfer learning for InceptionV3. The advantage of this transfer learning is that the pre-trained InceptionV3 model in the ImageNet dataset (Russakovsky et al., 2015) can be directly used to extract CNN features of small data samples (Pan and Yang, 2009). Therefore, InceptionV3 (Szegedy et al., 2016) was used in this study.

Fig. 5(a) illustrates schematic diagram of the InceptionV3 network. The function of this network is to transform the CNN features of the original image (resized 299 × 299 image) into discriminative features through feature dimension reduction and optimisation. Among them, the blocks of 3 × Inception, 5 × Inception and 2 × Inception increase the depth and width of the CNN network. In these 3 blocks, “Base” is the output of the former block, and “Filter Concat” is the synthesized output of the final CNN features. In this study, by using the 5 blocks in Fig. 5(a) the obtained CNN features were flattened into a 13,1072(8 × 8 × 2048)-dimensional vector as the input of the LSTM.

Fig. 5(b) illustrates schematic diagram of the LSTM network. LSTM can be considered as a special neuron with 4 inputs and 1 output. Where z, z_i, z_o and z_f are the control signal of LSTM. These 4 signals are input into the input gate, output gate and forget gate in order to obtain the output y^t. Memory units c^t and h^t generated in this process are brought into the next LSTM. It makes LSTM have a memory function (t = 1, 2,…, 30). The activation function g of z is the *tanh* function within the interval [−1, 1]. The activation function f of z_i, z_o and z_f is the *Sigmoid* function within the interval [0, 1]. The activation function h

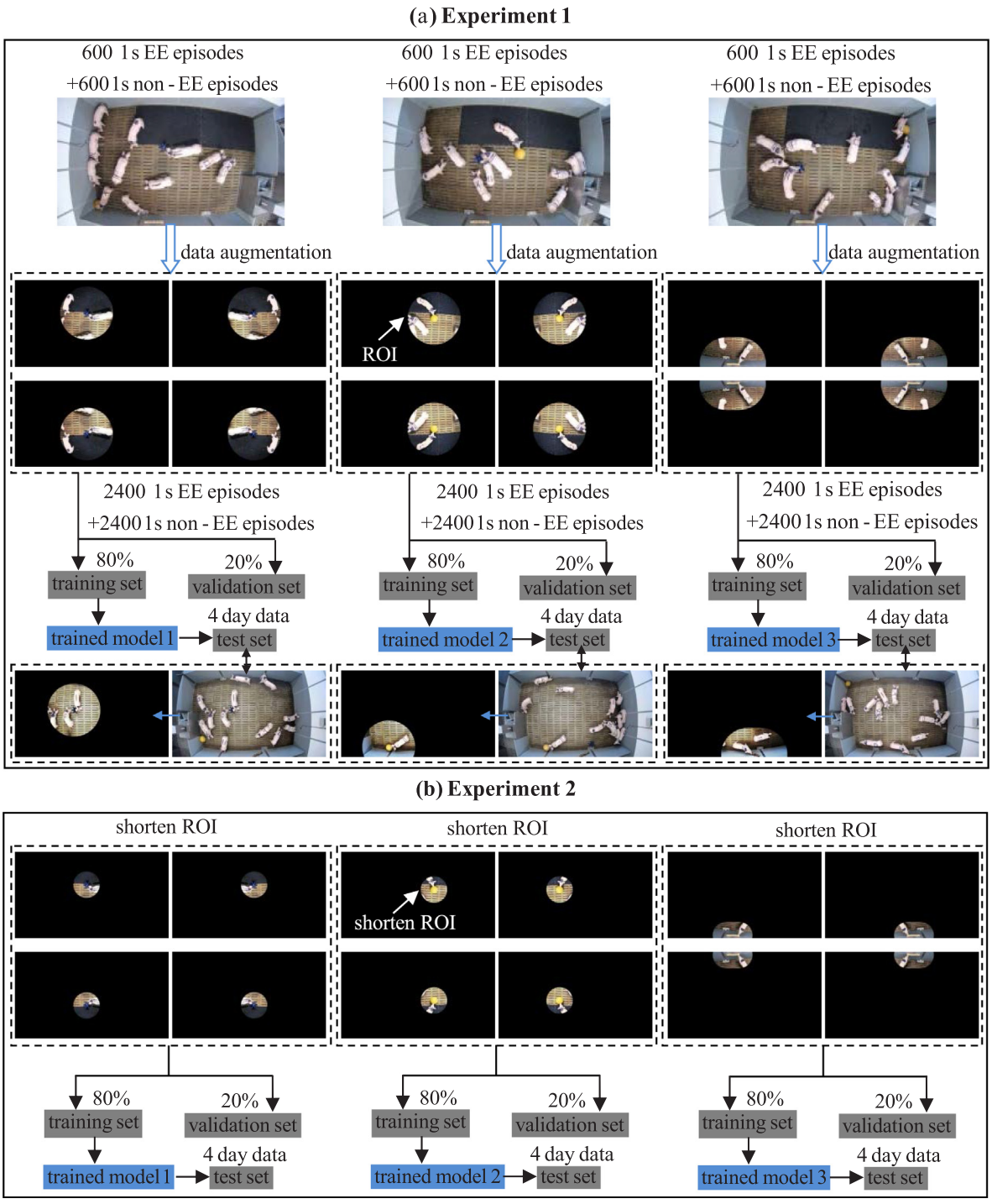

Fig. 3. Allocation of training set, validation set and test set.

Table 1
Description of enrichment engagement behaviour (EE) towards the three objects.

Behaviour category	Behaviour description
EE blue ball	One or multiple pigs actively engage with the blue ball by pushing the ball with the snout, by biting or chewing the protrusions of the ball, by carrying the ball or by throwing the ball. The event ends when the pig(s) shift its/their attention away from the ball.
EE golden ball	One or multiple pigs actively engage with the golden ball by pushing the ball with the snout, by kicking the ball with the front legs or by mounting the ball. If the ball moves and the pig(s) follow the ball, this is counted as part of the event. The event ends when the pig(s) shift its/their attention away from the ball.
EE wooden beam	One or multiple pigs actively engage with the wooden beam hanging from the pen inventory by turning the snout upwards to bite or push the beam or by lifting the beam with the head or the body. The event ends when the pig(s) shift its/their attention away from the beam.

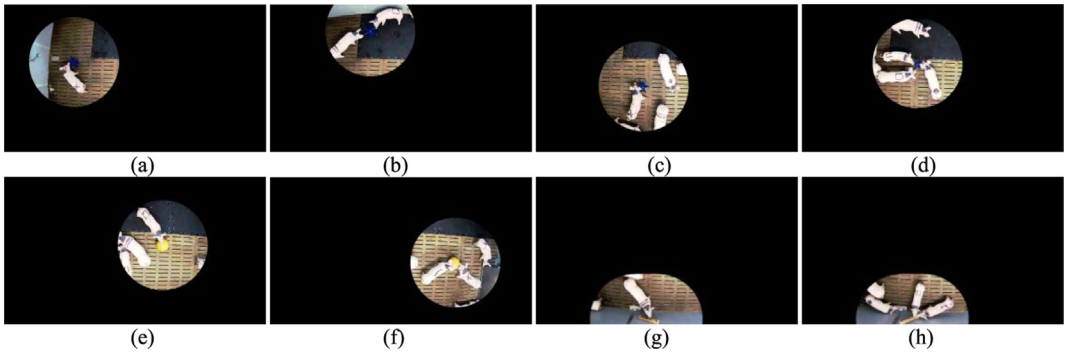

Fig. 4. The labelled data of enrichment engagement with the blue ball, golden ball and wooden beam: (a) chewing blue ball, (b) pushing blue ball, (c) carrying blue ball, (d) multiple pigs' engaging with blue ball, (e) pushing golden ball, (f) multiple pigs' engaging with golden ball, (g) pushing and biting wooden beam, and (h) multiple pigs' engaging with wooden beam.

of memory cell is the *tanh* function within the interval [−1, 1]. Eq. (1) was used to calculate c^t, h^t and y^t.

$$\begin{aligned} c^t = c' &= g(z)f(z_i) + cf(z_f) \\ h^t &= h(c') \\ y^t = a &= h(c')f(\boldsymbol{z_o}) \end{aligned} \tag{1}$$

Fig. 5(c) illustrates schematic diagram of the InceptionV3 and LSTM network from the vector view. Firstly, InceptionV3 was used to transform each frame of the video episode into a 131072-dimensional vector. In the first frame, this 131072-dimensional vector [x_1, x_2,..., x_{131072}] was multiplied with weights to obtain the control signals z, z_i, z_o and z_f, and then the output y^1 and the memory units c^1 and h^1 were obtained through LSTM. In the second frame, the corresponding another 131072-dimensional vector [x_1, x_2,..., x_{131072}] was multiplied with weights to obtain the control signals z, z_i, z_o and z_f, and then the output y^2 and the memory units c^2 and h^2 were obtained through LSTM. Among them, the memory units c^1 and h^1 in the first frame were brought into the second LSTM to determine the memory units c^2 and h^2 in the second frame. By using this method in turn, the 30-dimensional vector [y^1, y^2,..., y^{30}] corresponding to these 30 frames was used as total output of these 30 LSTM. This 30-dimensional vector was converted into a 2-dimensional vector through fully connected layer. Then, the *Softmax* function was used to convert all the elements of this 2-dimensional vector into values within the interval (0, 1) and normalise these values (the sum of all

type	patch size/stride or remarks	output size
conv block1	3×3/2	149×149×32
conv	3×3/1	147×147×32
conv padded	3×3/1	147×147×64
pool	3×3/2	73×73×64
conv block2	3×3/1	71×71×80
conv	3×3/2	35×35×192
conv	3×3/1	35×35×288
3×Inception block3	figure I	17×17×768
5×Inception block4	figure II	8×8×1280
2×Inception block5	figure III	8×8×2048

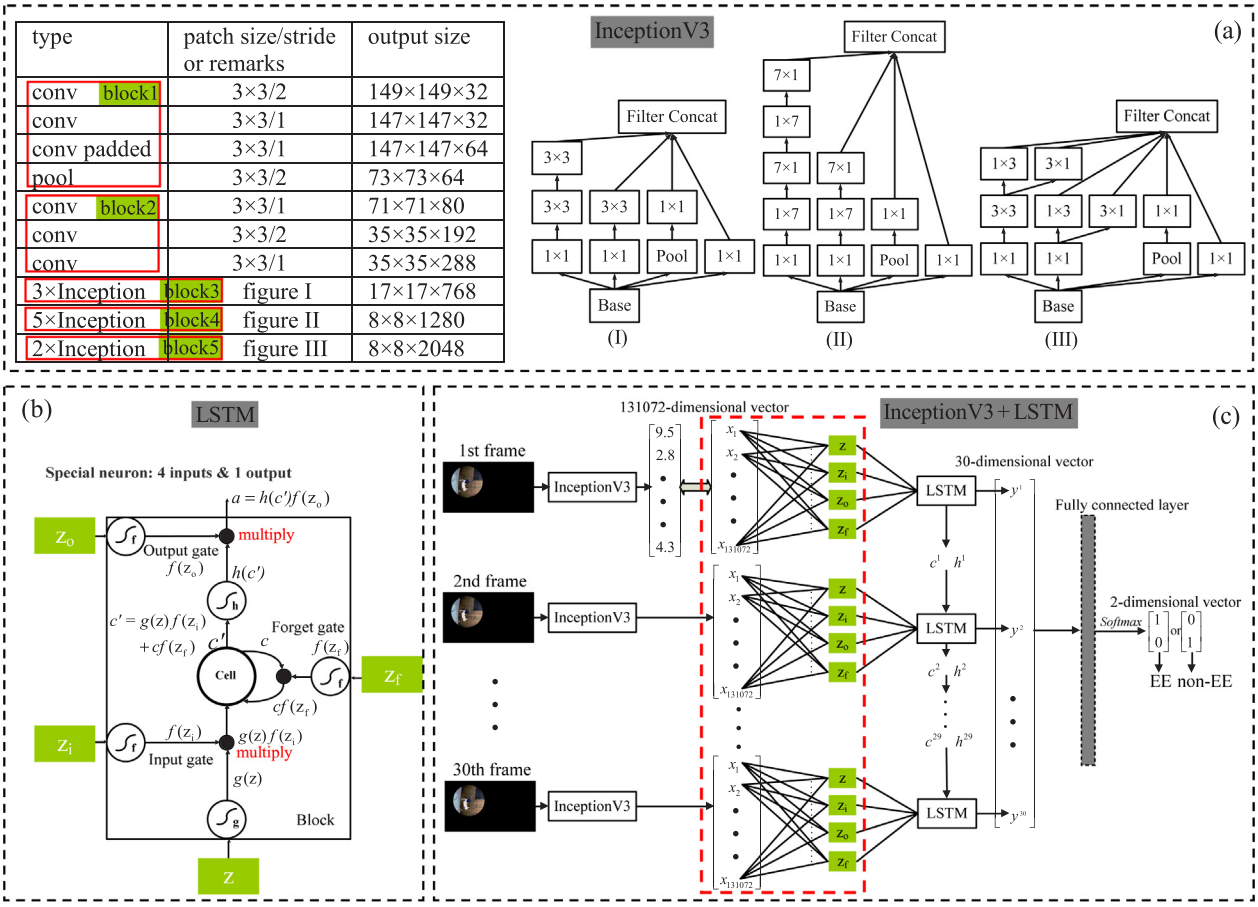

Fig. 5. Schematic diagram of the recognition of enrichment engagement with the blue ball, golden ball and wooden beam: (a) schematic diagram of InceptionV3, (b) schematic diagram of LSTM, and (c) connection manner of InceptionV3 and LSTM network from vector view.

values is 1). Finally, the class with the highest probability was selected as the predicted value 1 and another dimension as 0. Among them, the vector [1, 0] represents EE and the vector [0, 1] represents non-EE.

As the pigs show different motion patterns when engaging with the 3 different objects investigated in this study, the proposed algorithm was used to train 3 individual models to recognise engagement towards each of these 3 objects. Taking recognition of engagement with the blue ball in experiment 1 as an example, the specific steps of the training, validation and testing of the proposed algorithm are as follows:

1. LSTM randomly allocated training and validation sets. In the epoch, 80% of 1 s EE and 1 s non-EE episodes (i.e. 3840 1 s episodes) were randomly selected as training set, and the remaining 20% of episodes (i.e. 960 1 s episodes) were used as validation set.
2. The batchsize that represents the number of 1 s episodes input each time for training was set to 2. As a result, the training set was divided into 1920 (=3840/2) units for training and iterations in order to obtain the minimum loss. The model generated by these 1920 iterations was used to validate the 480 (=960/2) units in the validation set to obtain the accuracy. This process is termed an epoch.
3. After several epochs in turn, the loss of the model became smaller, while the accuracy improved. In the end, both the loss and the accuracy reached the optimal values.
4. The trained model was used to recognise EE episodes in the test set.

In order to evaluate the performance of the proposed algorithm, Eq. (2) was used to calculate accuracy, sensitivity and specificity.

$$\begin{aligned} Accuracy &= \frac{\text{Number of true positive and true negative episodes}}{\text{Total number of episodes}} \times 100\% \\ Sensitivity &= \frac{\text{Number of true positive episodes}}{\text{Number of true positive and false negative episodes}} \times 100\% \\ Specificity &= \frac{\text{Number of true negative episodes}}{\text{Number of false positive and true negative episodes}} \times 100\% \end{aligned} \tag{2}$$

where true positive episodes represent EE episodes that are classified as EE episodes. True negative episodes represent non-EE episodes that are classified as non-EE episodes. False positive episodes represent non-EE episodes that are classified as EE episodes. False negative episodes represent EE episodes that are classified as non-EE episodes.

Additionally, the cross-entropy function was used as the loss function in Eq. (3).

$$\text{loss} = -\sum_{c=1}^{M} y_c \log(p_c) \tag{3}$$

where c is the class, M is the number of all classes, y is the labelled result, and p is the predicted probability value normalised by the *Softmax* function. In this study, $M = 2$. Assuming that y is the vector [1, 0] representing EE, and p is the vector [0.92, 0.08]. As a result, the calculation process of loss is shown in Eq. (4).

$$\begin{aligned} \text{loss} &= -\sum_{c=1}^{2} y_c \log(p_c) = -y_1 \log(p_1) - y_2 \log(p_2) = -1 \log 0.92 - 0 \log 0.08 \\ &= 0.0362 \end{aligned} \tag{4}$$

when p is the vector [0.92, 0.08], the final predicted result of the *Softmax* function is [1, 0], which is the same as the labelled result y. This result indicates that EE episode is correctly classified as EE episode. Therefore, the number of true positive episodes will increase by 1.

3. Results and discussion

3.1. Enrichment engagement recognition

Fig. 6 illustrates tracking results of pigs engaged with the provided enrichment objects in the conditions of crowded pigs, dim illumination and dirty objects. The results indicate that the proposed tracking algorithm can be used to track EE in pigs in these conditions. The situation where golden or blue balls were moved into the feeder by pigs and then disappeared out of view was considered as non-EE event in this study. This is because due to the limited feeding space it becomes difficult for pigs to engage with the enrichment object in this position, and feeding pigs will quickly remove these objects from the feeder. Furthermore, this situation would not exist in pig pens equipped with other types of feeders, e.g. barrel feeders (Huang et al., 2018) and rectangular troughs (Yang et al., 2018), as these feeders cannot completely occlude the 2 balls.

Fig. 7 illustrates the discrimination change of CNN features of frames through the use of InceptionV3. 4-dimensional feature maps were manually selected from the output feature maps in each block. Through block1, the discrimination among these 4 feature maps is small, as these feature maps are all similar to the original image. From block2 to block5, the discrimination among these 4 feature maps gradually increased. The result indicates that InceptionV3 can be used to extract discriminative CNN features.

In order to further describe the relationship between these abstract CNN features and the motion patterns of EE behaviours, heat maps were generated and analysed. Heat maps can visualise the part of an image a CNN is focusing on. In other words, the heat map is mainly used to represent the position used for extracting features in the feature map, and the hot colours are used to visualise this position (Selvaraju et al., 2017). Fig. 8 illustrates the motion pattern difference of CNN features between EE and non-EE video sequences. By comparing the feature map and heat map of each video sequence, it can be seen that the abstract gray feature was converted into the specific position feature. It also can be seen from Fig. 8 that the CNN features of EE sequences change fast, while the CNN features of non-EE sequences almost remain unchanged. This difference occurred because there is continuous interaction and contact between pigs and objects during EE behaviour where the pigs are often moving either parts or the whole body, which is not the case for non-EE behaviours where no continuous interaction and contact between pigs and objects occur. The results indicate that the CNN features of EE and non-EE behaviours represent well the difference in motion patterns, which provides a basis for inputting CNN features into LSTM to classify EE and non-EE.

Table 2 illustrates the validation accuracy and validation loss when batchsize = 2, 4, 6 and 8, respectively and epoch = 200. When batchsize was increased from 2 to 8, it can be seen that the accuracy was reduced and the loss was increased. Therefore, the batchsize was set to 2 in this paper.

Fig. 9 illustrates the accuracy and loss curves of the proposed algorithm in the validation set. From Fig. 9(a-c), it can be seen that the proposed algorithm could recognise EE with the blue ball, golden ball and wooden beam with the accuracy of 95.2%, 95.4% and 97.3% and with the loss of 0.182, 0.186 and 0.125 in Experiment 1. By shortening the radius of the region of interest into a half of an average pig length (Fig. 9(d-f)), the accuracy of recognising EE with each of these 3 objects was improved to 96.9%, 97.1% and 97.9%, and the loss was reduced to 0.102, 0.134 and 0.058 in Experiment 2. This occurred because shortening the region of interest reduces the touching among pigs to a certain extent and thus reduces the interference of non-enrichment engaged pigs in the EE recognition. These results indicate that the proposed algorithm can be used to recognise EE behaviours of pigs, and halving the radius of the region of interest can improve the recognition accuracy and loss. Furthermore, shortening the region of interest may be of reference value for recognition of other behaviours of pigs. For instance, pigs' drinking movements are mainly concentrated on the mouth or the front half of the body. In the studies of the recognition of pig drinking, by reducing the region of interest from the entire pig body (Zhu et al., 2017) to the front half of the body or the head region (Zhang et al., 2019), the recognition accuracy may be improved.

Table 3 illustrates the number of true positive (TP), false negative (FN), false positive (FP) and true negative (TN) episodes and the accuracy, sensitivity and specificity of recognising EE with the blue ball,

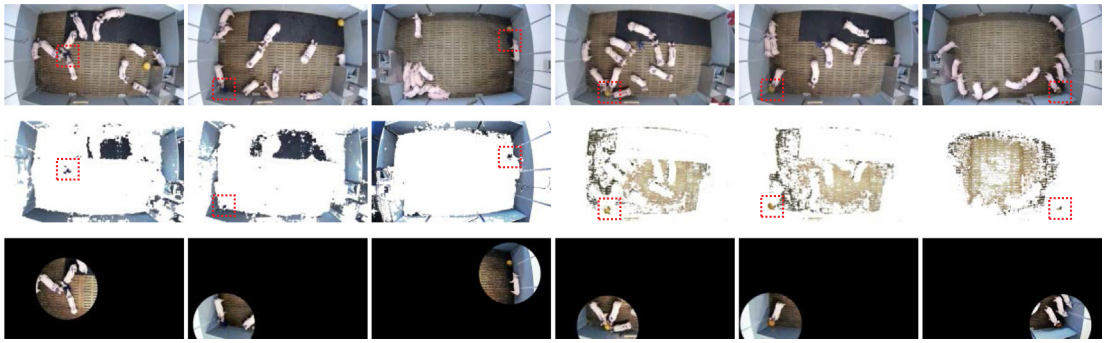

Fig. 6. Tracking results of enrichment engagement in pigs in the conditions of crowded pigs, dim illumination and dirty objects.

Fig. 7. Discrimination change of CNN features of frames through using InceptionV3.

golden ball and wooden beam in the test set. The accuracy, sensitivity and specificity of recognising EE with blue ball were 96.5%, 96.3% and 96.6%, respectively. The accuracy, sensitivity and specificity of recognising EE with golden ball were 96.8%, 96.2% and 96.8%, respectively. The accuracy, sensitivity and specificity of recognising EE with wooden beam were 97.6%, 97.8% and 97.6%, respectively. The results indicate that the proposed algorithm can be used to recognise EE behaviours of pigs, which is consistent with the results in Fig. 9.

Possible reasons for the false recognition of EE with the blue ball, golden ball and wooden beam are as follows:

For the blue ball, EE was falsely recognised as non-EE due to the pigs showing special behaviour variations of EE similar to other behaviours such as head-to-head aggression (0.5% of episodes), or due to the displacement of the blue ball by the pigs being very small (0.7% of episodes), or due to the pigs crowding around the blue ball (1.3% of episodes) or due to lens distortion (1.2% of episodes). Possible reasons for non-EE being falsely recognised as EE with the blue ball include pigs accidentally coming into contact with the blue ball during aggressive interactions (1.5% of episodes), during frightening events (1.2% of episodes) or while a pigs was exploring the pen including the floor (0.7% of episodes).

For the golden ball, EE was falsely recognised as non-EE due to that the displacement of the golden ball created by the pigs being very small (1.1% of episodes) or due to the pigs being frightened (0.8% of episodes) or attacked (1.9% of episodes) by other pigs during the engagement with the golden ball. Possible reasons for non-EE to be falsely recognised as EE with the golden ball include pigs accidentally coming into contact with the golden ball during aggressive interactions (1.0% of episodes), during frightening events (0.4% of episodes), during walking (0.5% of episodes) or while a pig was exploring the pen including the floor (1.3% of episodes).

For the wooden beam, EE was falsely recognised as non-EE due to the engagement duration with the wooden beam being very short (0.5% of episodes) or due to the pigs being frightened (0.6% of episodes) or attacked (1.1% of episodes) by other pigs during the engagement with the wooden beam. Reasons for non-EE to be falsely recognised as EE with the wooden beam include pigs accidentally coming into contact with the wooden beam during aggressive interactions (0.8% of

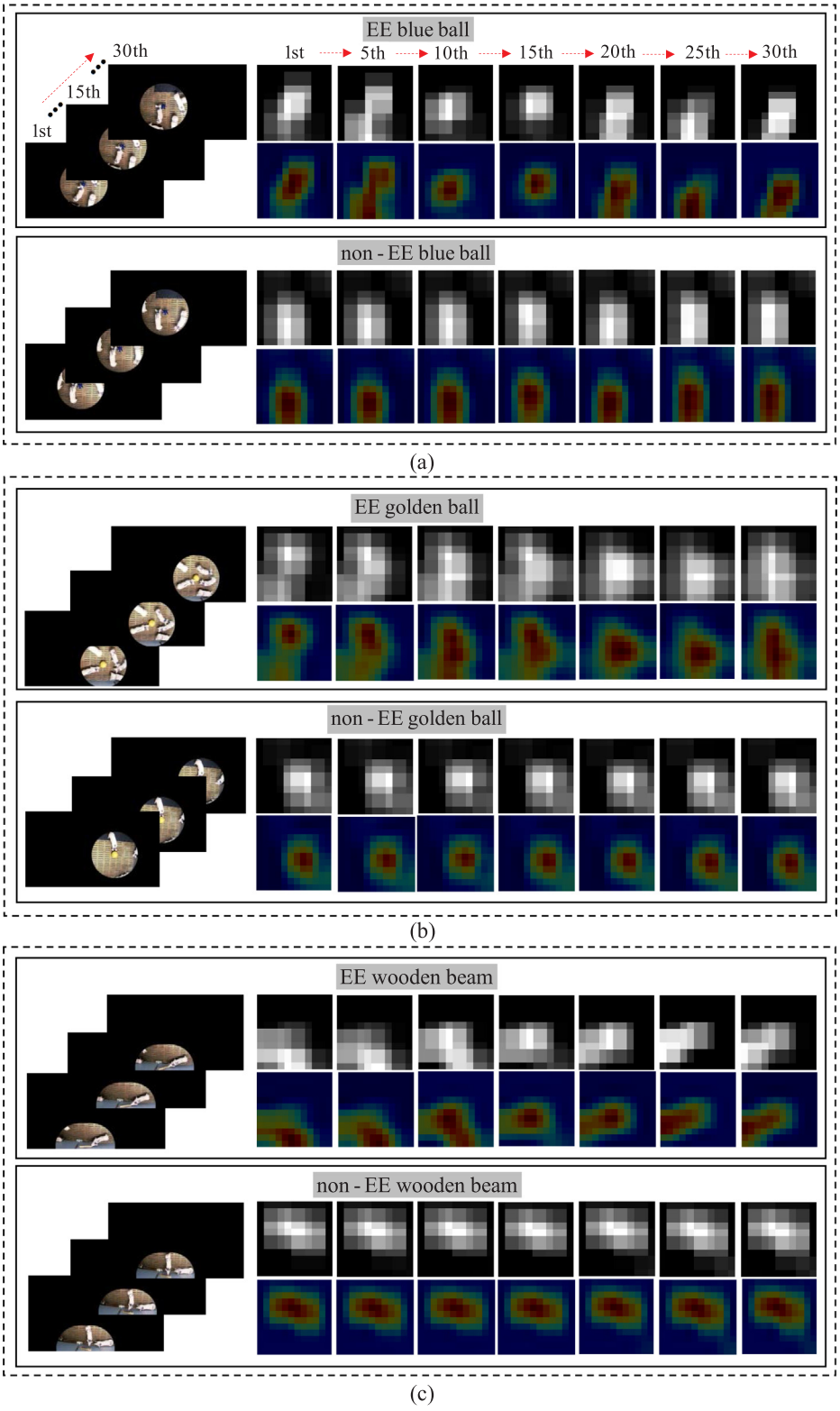

Fig. 8. Motion pattern difference of CNN features between enrichment engagement and non-enrichment engagement video sequences.

episodes), during walking (0.5% of episodes) or during crowding of pigs around the wooden beam (1.1% of episodes).

3.2. Pigs preference determination

In order to also investigate the preference of pigs for different objects, Fig. 10 shows the duration of the recognised and the labelled EE for each of the three objects during each of the 8 h of the 4 days included in the test set. Within each hour, the length of colour bar of the recognised EE was close to the length of colour bar of the labelled EE. This indicates that the recognition results of the proposed algorithm are consistent with the labelling results. In the 4 days of 32-hour data, the duration of the labelled EE with the blue ball, golden ball and wooden beam was 27330, 2151 and 6461 s, respectively, and the duration of the

Table 2
Validation accuracy and validation loss when batchzise = 2, 4, 6 and 8, respectively and epoch = 200.

Behaviour category	EE blue ball		EE golden ball		EE wooden beam	
Batchsize (Epoch = 200)	Validation accuracy	Validation loss	Validation accuracy	Validation loss	Validation accuracy	Validation loss
2	95.2%	0.182	95.4%	0.186	97.3%	0.125
4	94.2%	0.218	94.2%	0.347	96.4%	0.146
6	90.8%	0.450	92.1%	0.482	94.6%	0.212
8	89.4%	0.488	88.8%	0.480	91.2%	0.283

recognised EE was 26323, 2070 and 6317 s, respectively. Therefore, the blue ball accounts for approximate 76% of the EE, the wooden beam accounts for approximate 18% of the EE and the golden ball accounts for 6.0% of the EE, for both the labelled and recognised durations. Based on the duration of engagement only, pigs seem to prefer the blue ball with protrusions over the hanging wooden beam and the golden ball. The quality of an object from the pigs' point of view can depend on many factors including the material, the shape, the location within the pen, the reach ability of the object, how easily the object get soiled, the number of ways it can be manipulated, the novelty in the object and whether it promotes social facilitation. The blue ball is made of a chewable material and it has a shape that gives unpredictable movement when manipulated and provides many manipulation options including chewing, shaking, throwing and pushing (Larsen et al., 2019). Further, the protrusions give a low contact-surface decreasing soiling and ensure that multiple pigs can manipulate it at the same time. The blue ball also makes it possible for the pigs to manipulate it by their natural rooting movements of the neck, head and snout (e.g. foraging behaviour). This may be some of the reason why the blue ball was used for longer time than the other objects investigated.

The same conclusion on object preference was obtained by the developed algorithm and thereby the laborious manual labelling shows the possibility for the algorithm to replace the manual labelling process. However, the algorithm is developed to detect the object and thus, it lacks information about the number of pigs engaged with the object. Often, multiple pigs were engaged with the blue ball which was not as

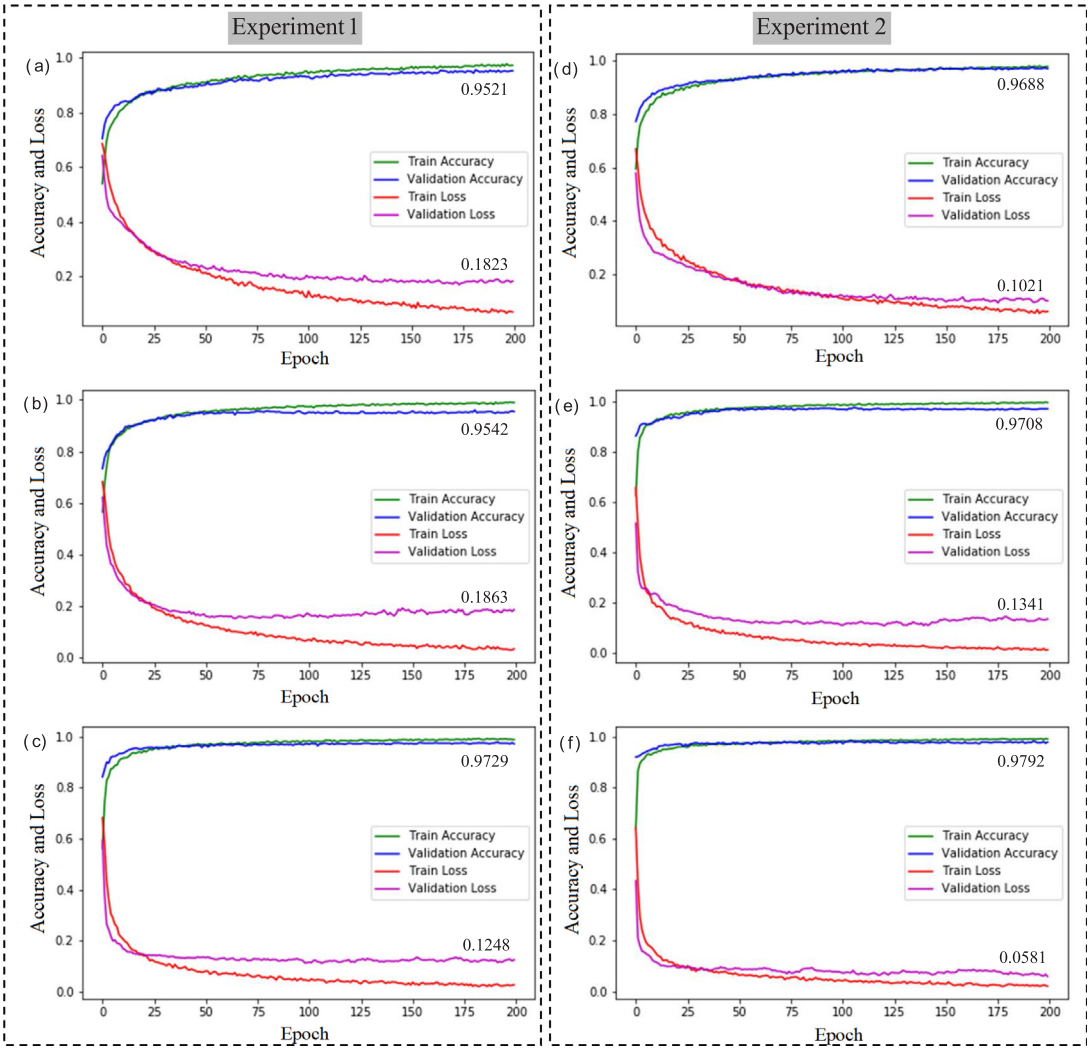

Fig. 9. Accuracy and loss curves of the proposed algorithm in the validation set: (a-c) accuracy and loss curves of recognising enrichment engagement with the blue ball, golden ball and wooden beam in Experiment 1, and (d-f) accuracy and loss curves of recognising enrichment engagement with the blue ball, golden ball and wooden beam after shortening EE region of interest in Experiment 2.

Table 3
The number of true positive (TP), false negative (FN), false positive (FP) and true negative (TN) episodes and the accuracy, sensitivity and specificity of recognising enrichment engagement (EE) with the blue ball, golden ball and wooden beam in the test set.

Behaviour category	TP	FN	FP	TN	Accuracy	Sensitivity	Specificity
EE blue ball	26,323	1007	3025	84,845	96.5%	96.3%	96.6%
EE golden ball	2070	81	3606	109,443	96.8%	96.2%	96.8%
EE wooden beam	6317	144	2621	106,118	97.6%	97.8%	97.6%

often the case for the two other objects. If this was taken into consideration, the concluded preference of the blue ball would have been even stronger. As the tracking algorithm proposed in this paper located the EE region of interest and thus the body of EE pigs in this region is entire while that of non-EE pigs is not. Based on the area difference between EE and non-EE pig individuals, in future work the threshold of the area of each pig could be set to distinguish EE and non-EE pigs and then count the number of EE pigs by using a connection domain-based image processing algorithm or a deep learning-based pig detection algorithm. Another important parameter is how many individual pigs in the pen the object engages, although that would demand individual recognition of the pigs. Thus, to provide the full picture of EE, the algorithm still has room for further development.

3.3. Discussion on advantages and disadvantages of the methodology

This paper trained 3 models for recognition of EE with the blue ball, golden ball and wooden beam by combining InceptionV3 and LSTM and then recognised these 3 EE behaviours from a series of non-EE behaviours (i.e. lying, walking, drinking, feeding, running, chasing and aggression). At the same time, these 3 EE behaviours were also classified by using the proposed algorithm. Among them, object detection is the key to the proposed tracking algorithm. In the location of the object's centroid, this paper mainly adopts a HSV colour space transformation and connected domain-based image processing method. The coordinate of the obtained centroid has a certain deviation, while it does not affect the location of the EE region of interest. On the other hand, the deep learning-based object detection algorithms (e.g. Faster R-CNN (Girshick, 2015), SSD (Liu et al., 2016) and YOLO (Redmon et al., 2016)) can also be used to locate these 3 objects. However, the occluded part of objects still cannot be detected, and thus the deviation of the centroid of objects still exists. As the deep learning-based object detection methods need to perform labelling of large amounts of data as well as training and testing of models. The advantage of the proposed tracking algorithm is that the calculation amount is small, and it can locate the centroid of objects more directly.

As each colour corresponds to a special range of H, S and V components (Gonzalez and Woods, 2007), the proposed tracking algorithm has the potential to locate objects with other specific colours in order to recognise pig EE behaviours towards the objects with different colours, shapes and materials and to study pigs' preference for other objects than the ones investigated in the current study. On the other hand, only the duration of EE in the first 4 days after fattening was counted in this study, and then the pigs' preference for objects was preliminarily determined. In future studies of pigs' preference for objects, the proposed algorithm could also to be used to investigate EE during longer periods, for more pens and for pigs at different growth stages.

In the previous study of aggressive behaviours of pigs (Chen et al., 2020), the entire image in the video was used as data for training the models, which belongs to group level analysis. This study further located the EE pig individuals and only used the image in the region of interest around the object as data for training the models. This difference depends on behavioural characteristics between aggression and EE. Namely, the motion velocity and interaction pattern of aggressive behaviours between adjacent frames changes much faster than that of non-aggressive behaviours, while the motion velocity and interaction pattern of EE behaviours is similar to that of other behaviours. Therefore, the InceptionV3 and LSTM network proposed in this study has reference value for classifying behaviours with similar motion patterns.

4. Conclusion

This paper proposed a recurrent neural network-based deep learning algorithm by combining InceptionV3 and LSTM to automatically recognise EE of pigs and preliminarily determine their preference to different objects. The proposed tracking algorithm can be used to track the objects with different colours, shapes and materials in the conditions of crowded pigs, dim illumination and dirty objects. InceptionV3 can be used to extract the discriminative CNN features. The LSTM spatial-temporal features can be used to distinguish the

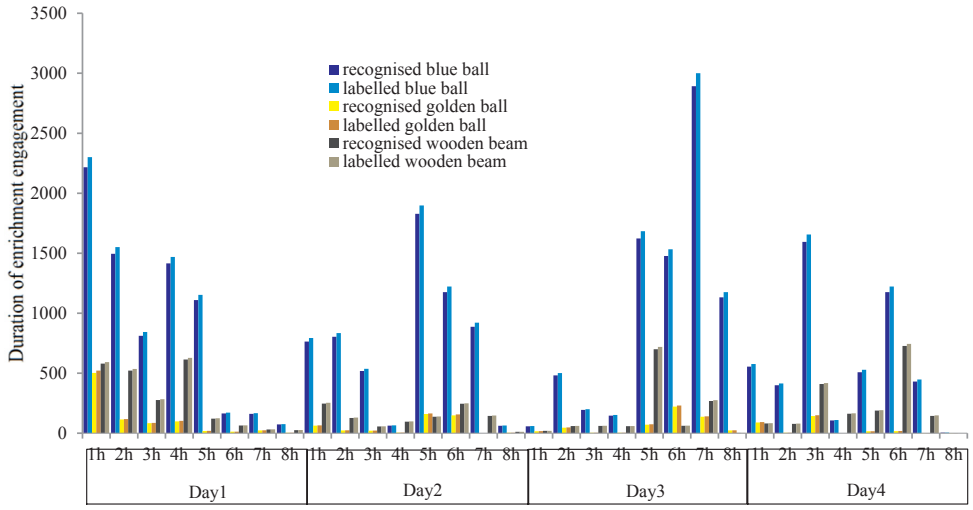

Fig. 10. Comparison of the recognised and the labelled duration of enrichment engagement with the blue ball, golden ball and wooden beam in each of the 8 h during the 4 days after fattening (the test set).

motion pattern difference of EE and non-EE. In validation set, the proposed algorithm could recognise EE with the blue ball, golden ball and wooden beam with an accuracy of 95.2%, 95.4% and 97.3%, respectively. By shortening the radius of the region of interest into a half of average length of pig body, the corresponding accuracy could be further improved into 96.9%, 97.1% and 97.9%, respectively. In test set, the proposed algorithm could recognise EE with blue ball with an accuracy of 96.5%, a sensitivity of 96.3% and specificity of 96.6%, recognise EE with golden ball with an accuracy of 96.8%, a sensitivity of 96.2% and specificity of 96.8%, and recognise EE with wooden beam with an accuracy of 97.6%, a sensitivity of 97.8% and specificity of 97.6%. The proportion of EE with the blue ball, golden ball and wooden beam was 75.8%, 6.0% and 18.2%, respectively. The results indicate that the proposed method can be used to recognise pig EE behaviours, and halving the radius of the region of interest can improve the recognition accuracy of EE. Moreover, the preference of pigs to objects based on the engagement duration only is blue ball > wooden beam > golden ball. This study recognised pigs' EE behaviours and preliminarily determined their preference to different objects, which has practical application value. Furthermore, the proposed InceptionV3 and LSTM network has reference value for classifying behaviours with similar motion patterns.

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.

Acknowledgements

This work was a part of the project funded by the "National Natural Science Foundation of China", China (grant number: 31872399). This work was created within a research project of the Austrian Competence Centre for Feed and Food Quality, Safety and Innovation (FFoQSI), Austria (grant number: 854182). The COMET-K1 competence centre FFoQSI is funded by the Austrian ministries BMVIT, BMDW and the Austrian provinces Niederoesterreich, Upper Austria and Vienna within the scope of COMET-Competence Centers for Excellent Technologies. The programme COMET is handled by the Austrian Research Promotion Agency FFG. This work was also a part of the "Auto Play Pig" project funded by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement (grant number: 842555) and the "China Scholarship Council (CSC)", China (File No. 201808320269). The authors would like to thank Barbara Metzler-Zebeli, Julia Vötterl, Jutamat Klinsoda and Thomas Enzinger from the Institute of Animal Nutrition and Functional Plant Compounds of University of Veterinary Medicine Vienna for excellent cooperation and practical support as well as Dong Liu for his help in the experiment.

References

Brown, S.M., Peters, R., Nevison, I.M., Lawrence, A.B., 2018. Playful pigs: evidence of consistency and change in play depending on litter and developmental stage. Appl. Animal Behav. Sci. 198, 36–43.

Banerjee, I., Ling, Y., Chen, M.C., Hasan, S.A., Langlotz, C.P., Moradzadeh, N., Chapman, B., Amrhein, T., Mong, D., Rubin, D.L., Farri, O., Lungren, M.P., 2019. Comparative effectiveness of convolutional neural network (CNN) and recurrent neural network (RNN) architectures for radiology text report classification. Artif. Intell. Med. 97, 79–88.

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

Donahue, J., Hendricks, L.A., Guadarrama, S., Rohrbach, M., Venugopalan, S., Saenko, K., Darrell, T., 2015. Long-term recurrent convolutional networks for visual recognition and description. IEEE Conf. Comp. Vis. Pattern Recogn. 2625–2634.

Fu, L., Zhou, B., Li, H., Schinckel, A.P., Liang, T., Chu, Q., Li, Y., Xu, F., 2018. Teeth clipping, tail docking and toy enrichment affect physiological indicators, behaviour and lesions of weaned pigs after re-location and mixing. Livestock Sci. 212, 137–142.

Gonzalez, R., Woods, R., 2007. Digital Image Processing, third ed. Prentice-Hall.

Girshick, R. (2015). Fast R-CNN. The IEEE International Conference on Computer Vision (ICCV), 1440–1448.

Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Comput. 9 (8), 1735–1780.

Huang, W., Zhu, W., Ma, C., Guo, Y., Chen, C., 2018. Identification of group-housed pigs based on Gabor and Local Binary Pattern features. Biosyst. Eng. 166, 90–100.

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.Y., & Berg, A.C. (2016). SSD: Single Shot MultiBox Detector. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 8-16 October 2016; Springer: Berlin/Heidelberg, Germany, 2016; pp. 21–37.

Lahrmann, H.P., Hansen, C.F., ÓEath, R.B., Busch, M.E., Nielsen, J.P., Forkman, B., 2018. Early intervention with enrichment can prevent tail biting outbreaks in weaner pigs. Livestock Sci. 214, 272–277.

Larsen, M.L.V., Jensen, M.B., Pedersen, L.J., 2019. Increasing the number of wooden beams from two to four increases the exploratory behaviour of finisher pigs. Appl. Animal Behav. Sci. 216, 6–14.

Nasirahmadi, A., Edwards, S.A., Sturm, B., 2017. Implementation of machine vision for detecting behaviour of cattle and pigs. Livestock Sci. 202, 25–38.

Nguyen, H.T., Nguyen, C.T., Bao, P.T., Nakagawa, M., 2018. A database of unconstrained Vietnamese online handwriting and recognition experiments by recurrent neural networks. Pattern Recogn. 78, 291–306.

Pan, S.J., Yang, Q., 2009. A survey on transfer learning. IEEE Trans. Knowl. Data Eng. 22 (10), 1345–1359.

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

Redmon, J., Divvala, S., Girshick, R., Farhadi, A., 2016. You only look once: unified, real-time object detection. IEEE Conf. Comp. Vision Pattern Recog. (CVPR) 779–788.

Srivastava, N., Mansimov, E., Salakhutdinov, R., 2015. Unsupervised learning of video representations using LSTMs. International Conference on Machine Learning.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z., 2016. Rethinking the inception architecture for computer vision. The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2818–2826.

Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D., 2017. Grad-CAM: visual explanations from deep networks via gradient-based localization. IEEE Int. Conf. Comp. Vis. (ICCV) 618–626.

Turner, S.P., Farnworth, M.J., White, I.M.S., Brotherstone, S., Mendl, M., Knap, P., Penny, P., Lawrence, A.B., 2006. The accumulation of skin lesions and their use as a predictor of individual aggressiveness in pigs. Appl. Animal Behav. Sci. 96 (3), 245–259.

Telkänranta, H., Swan, K., Hirvonen, H., Valros, A., 2014. Chewable materials before weaning reduce tail bitingin growing pigs. Appl. Animal Behav. Sci. 157, 14–22.

Tsironi, E., Barros, P., Weber, C., Wermter, S., 2017. An analysis of convolutional long short-term memory recurrent neural networks for gesture recognition. Neurocomputing 268, 76–86.

Tian, M., Guo, H., Chen, H., Wang, Q., Long, C., Ma, Y., 2019. Automated pig counting using deep learning. Comput. Electron. Agric. 163, 104840.

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

Yang, A., Huang, H., Zheng, B., Li, S., Gan, H., Chen, C., Yang, X., Xue, Y., 2020. An automatic recognition framework for sow daily behaviours based on motion and image analyses. Biosyst. Eng. 192, 56–71.

Zhu, W., Guo, Y., Jiao, P., Ma, C., Chen, C., 2017. Recognition and drinking behaviour analysis of individual pigs based on machine vision. Livestock Sci. 205, 129–136.

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.

Zhang, Y., Cai, J., Xiao, D., Li, Z., Xiong, B., 2019. Real-time sow behavior detection based on deep learning. Comput. Electron. Agric. 163, 104884.

Zhu, X., Chen, C., Zheng, B., Yang, X., Gan, H., Zheng, C., Yang, A., Mao, L., Xue, Y., 2020. Automatic recognition of lactating sow postures by refined two-stream RGB-D faster R-CNN. Biosyst. Eng. 189, 116–132.