Computers and Electronics in Agriculture 174 (2020) 105391

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

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

Automatically detecting pig position and posture by 2D camera imaging and deep learning

Martin Riekert[a,*], Achim Klein[a], Felix Adrion[b], Christa Hoffmann[c], Eva Gallmann[d]

[a] Chair of Information Systems 2, University of Hohenheim, Schwerzstraße 35, 70599 Stuttgart, Germany
[b] Research Division on Competitiveness and System Evaluation, Agroscope, Tänikon 1, 8356 Ettenhausen, Switzerland
[c] Bildungs- und Wissenszentrum Boxberg, Seehöfer Straße 50, 97944 Boxberg-Windischbuch, Germany
[d] Institute of Agricultural Engineering, University of Hohenheim, Garbenstraße 9, 70599 Stuttgart, Germany

ARTICLE INFO

Keywords:
Deep learning
Object detection
Pig position and posture detection
Pig behavior
2D vision cameras

ABSTRACT

Prior livestock research provides evidence for the importance of accurate detection of pig positions and postures for better understanding animal welfare. Position and posture detection can be accomplished by machine vision systems. However, current machine vision systems require rigid setups of fixed vertical lighting, vertical top-view camera perspectives or complex camera systems, which hinder their adoption in practice. Moreover, existing detection systems focus on specific pen contexts and may be difficult to apply in other livestock facilities. Our main contribution is twofold: First, we design a deep learning system for position and posture detection that only requires standard 2D camera imaging with no adaptations to the application setting. This deep learning system applies the state-of-the-art Faster R-CNN object detection pipeline and the state-of-the-art Neural Architecture Search (NAS) base network for feature extraction. Second, we provide a labelled open access dataset with 7277 human-made annotations from 21 standard 2D cameras, covering 31 different one-hour long video recordings and 18 different pens to train and test the approach under realistic conditions. On unseen pens under similar experimental conditions with sufficient similar training images of pig fattening, the deep learning system detects pig position with an Average Precision (AP) of 87.4%, and pig position and posture with a mean Average Precision (mAP) of 80.2%. Given different and more difficult experimental conditions of pig rearing with no or little similar images in the training set, an AP of over 67.7% was achieved for position detection. However, detecting the position and posture achieved a mAP between 44.8% and 58.8% only. Furthermore, we demonstrate exemplary applications that can aid pen design by visualizing where pigs are lying and how their lying behavior changes through the day. Finally, we contribute open data that can be used for further studies, replication, and pig position detection applications.

1. Introduction

Understanding pig behavior is important due to the impact of pig behavior on pig well-being (Vranken and Berckmans, 2017). For instance, reduced activity may help to uncover diseases or discomfort in animals (Matthews et al., 2016; Weary et al., 2009). Behavioral changes can be detected early on at the individual or group level with appropriate sensor data (Madsen and Kristensen, 2005; Maselyne et al., 2017). For this purpose, many inexpensive commercial surveillance cameras with high resolution are installed in research facilities. Human experts spend days for manually annotating positions and postures in videos (Brendle and Hoy, 2011). The annotation task is tedious, cannot be scaled, and thus can only be conducted in research settings (Kalbe et al., 2018). One promising approach for automating this task is the use of RFID sensors, which are inexpensive and allow individual identification of pigs (Adrion et al., 2018; Hammer et al., 2016, 2017). This approach seems reliable for individual hotspot monitoring, but it is necessary to install RFID antennas at every location of interest in the housing environment (hotspot) (Kapun et al., 2018), and equip each animal with one or two RFID ear tags (Maselyne et al., 2016). Additionally, the posture of a pig is not detectable in a standard RFID setup, but only the presence of the animal in the antenna field (Adrion et al., 2018). Alternative approaches are posture detection with accelerometers, which are frequently used in dairy cows (Borchers et al., 2016) and automatic position detection systems (Nasirahmadi et al., 2017b). However, using image analysis for behavior monitoring is preferable in pigs, because only comparatively small sensors can be used and the sensors are always at risk of destruction due to the

[*] Corresponding author.
E-mail address: martin.riekert@uni-hohenheim.de (M. Riekert).

https://doi.org/10.1016/j.compag.2020.105391
Received 10 December 2018; Received in revised form 26 March 2020; Accepted 28 March 2020

exploration behavior of the animals.

Hence, automatic position detection systems for pigs are important for continuous monitoring without manual work. Such systems require automatic object detection in video images, which is the algorithmic task of localizing, and classifying objects in images. Most current approaches binarize an image into black and white pixels, remove pixels that are too small, and then try to fit ellipses around white pixels (McFarlane and Schofield, 1995). This approach achieves an accuracy between 88.7% (Kashiha et al., 2013) and 95.8% (Nasirahmadi et al., 2015). Modifications to ellipse-based approaches detect the position of pigs without enhancing lighting or other methods to improve the visibility of pigs (Brünger et al., 2018; Guo et al., 2015). One of the adaptations is to use a manually labeled starting point, which is then used for the detection in the next frame (Brünger et al., 2018). Errors occur if pigs lie close together or during mounting behavior. This detection system achieves on one pen with one top view camera a detection rate of 90% (Brünger et al., 2018). However, such ellipse-based approaches require a controlled experimental environment with consistent lighting conditions, in some cases manual blacking of out of pen equipment from the video, and a vertically aligned camera that points exactly downwards (top view). Furthermore, most of these approaches identify pigs of approximately the size in which pigs are shown in the experimental videos. Therefore, it is difficult to generalize these approaches to other settings.

Pig locomotion detects whether a pig is changing its location, but does not detect the posture of the pig. Pig locomotion can achieve an accuracy of up to 89.8% (Kashiha et al., 2014). The approach proposed by Ahrendt et al. (2011) integrates geometric distance between the centroid of the pig in the previous video frame as well as differences in the color of pixels. This approach requires a starting point and a high discrepancy between the color value of the pig and the floor; however, these requirements are not guaranteed in settings with direct daylight shining on the floor from a window. The performance of the algorithm was evaluated in a one-time observation, in which one pig was successfully tracked over 8 min.

Another approach is to mark pigs with different colors and to detect the colors in the image (Navarro Jover et al., 2009). This approach allows to identify each pig individually. While the success rate of the algorithm is reported as 64.6%, the approach requires a high amount of work due to the requirement to paint each pig in a different color. Another approach relies on manually developed features and uses logistic regression with elastic net regularization. The procedure can distinguish straw from pigs. However, experiments under daylight (without artificial light) were excluded because the approach could not handle such situations (Nilsson et al., 2015).

Pig lying behavior is detected by an ellipsoid-based approach using floor-viewing cameras mounted on ceilings to detect the position of pigs (Nasirahmadi et al., 2015). They achieve an accuracy of 95.8%. The approach to locate pigs in videos is also used to study mounting behavior (Nasirahmadi et al., 2016). After detecting pig segments by ellipsoids, the ellipsoids are further processed to detect the spatial orientation of pigs and the mounting behavior is predicted by spotting ellipses that are merging with other ellipses.

Other approaches use 3D Kinect cameras to estimate the weight of pigs (Condotta et al., 2018; Kongsro, 2014; Pezzuolo et al., 2018), or to identify standing pigs by detecting objects that are higher than the floor (Kim et al., 2017; Matthews et al., 2017). 3D cameras also detect the tail height and are used to detect tail biting (D'Eath et al., 2018). However, the Kinect depth sensor has a rather low maximum distance range of 4.5 m (Kim et al., 2017; Mallick et al., 2014). Furthermore, the resolution of the 3D video is limited to 512 × 424 pixels; hence, only a small area of a pen can be observed with one camera.

Machine vision research also contributed to automatically detecting objects in video images (Everingham et al., 2010). The most effective object detection algorithms adopt deep learning (Huang et al., 2016; LeCun et al., 2015). These deep learning algorithms have been used together with 3D cameras to study sow lying behavior evaluated on test data of 1 sow (Zheng et al., 2018) and 2D cameras to detect the feeding behavior for group housed pigs evaluated on test data of 1 pen with 4 pigs (Yang et al., 2018).

Overall, our discussion of prior research results highlights an important gap in the literature, i.e., approaches for automatically detecting pigs' positions and postures using standard 2D cameras in real-world settings under different lighting conditions and camera directions. Additionally, it is still unknown for researchers and practitioners in pig farming how to adopt and configure deep learning techniques to achieve high mAP and what the limitations of deep learning algorithms are (Kamilaris and Prenafeta-Boldú, 2018; Nasirahmadi et al., 2017a). Overcoming these limitations would help foster the applicability of pig position detection systems in real-world settings. We propose and evaluate a solution specifically for detecting pigs in a large number of different pens including the lying behavior using realistic camera and lighting settings. Our research contributes to the applicability of pig detection systems by transferring and applying state-of-the-art deep learning-algorithms from machine vision to automatically extract the position of pigs within their pen and determine the lying and not-lying behavior of pigs. The contributions are (1) an image dataset with 7277 human-made annotations with respect to lying and not lying behavior of pigs, (2) parametrizations of deep learning algorithms to accurately extract the position and posture of pigs in still images from videos, and (3) an out of sample evaluation of our detection models under different pen setups using state-of-the-art object detection metrics.

2. Material and methods

2.1. Animals and test facility

Place of study was the Boxberg Teaching and Research Centre – Centre for pig rearing and pig breeding, which is a subunit of the Ministry of Rural Affairs and Consumer Protection of the federal state of Baden-Württemberg in Germany. Our study was conducted within these facilities in two different conventional housing systems for piglet rearing from about 5 kg to 30 kg and fattening pigs from about 30 kg to 117 kg. Major differences between the two housing systems were the pen design (e.g., floor structure) and number of pigs per pen. The distance from pen to window (day light) was different as well. All pigs studied have the same genetic background (German Genetic) and were born in the same conventional piglet production at the research centre.

In the Boxberg Teaching and Research Centre a pen is identified by an identifier (e.g., "B102"), which refers to the building as an alphabetical character (e.g., "B") and the compartment were the pen in the building is located (e.g., "1"). That is, the compartment identifier comprises the first two alphanumerical characters (e.g., "B1") and finally two numerical letters, which identify the specific pen in the compartment (e.g., "02").

The video recordings in our study were not specifically set up for our study or with machine vision applications in mind. The videos were recorded by standard 2D video cameras based on requirements of third party animal researchers to study pig activity. The third party animal researchers installed the cameras for human annotation of pig behavior including their activity level and other animal welfare parameters. Therefore, our approach should be well suited for other pig video recordings.

2.2. Experimental procedure and approach

Our approach treats detecting the position and posture of pigs as an object detection problem. Object detection is concerned with the task of providing algorithms for (1) localizing, and (2) classifying objects in images (Everingham et al., 2010). The localization task was defined by encompassing a pig with a rectangle (Everingham et al., 2010). This rectangle is termed a bounding box and defined as a rectangle in an

Table 1
Descriptive statistics of the dataset with respect to different pens. The abbreviations are: N: sample size, M: mean, SD: standard deviation, MIN: minimal value, Q1: first quartile, Q2: second quartile/median, Q3: third quartile and MAX: maximum value.

2.3.1 Pig compartments	Type		Pig fattening					Piglet rearing					
	Compartment identification		B1		C1		C2	D6					A1
	Number of unique pens in videos		4		6		5	2					1
	Number of housed pigs per pen		28		18		18	48					24
	Shape of pen area		4.25 m x 7.8 m		4.25 m x 5.2 m		4.25 m x 5.2 m	2.85 m x 7.95 m					2.99 m x 4.99 m
	Area in m^2 per pig (without pen equipment)		1.14		1.18		1.19	0.45					0.58
2.3.2 Camera	Type		HIKVision DS -2CD2125FWD-I IP Dome camera 2MP Full HD Outdoor 2.8 mm										Edimax IC-9110W
	Resolution		1280x720										640x480
	Frames per second (fps)		10										10
	Perspective		High angle shot (Figure 1 and Figure 2)					High angle shot (Figure 3)				Top view (Figure 4)	Top view (Figure 5)
	Pens visible in camera		2	1	1		1	2 (front camera)		2 (back camera)		2	1
	Complete pen visible in camera's images		No	No	Yes		Yes	No		No		No	No
	Number of unique cameras		2	3	6		5	1		1		2	1
2.3.3 Video image sample	Purpose		train	train	train	test	test	train	test	train	test	train	test
	Number of recordings		2	3	6	5	5	1	2	1	2	2	2
	Number of images		20	30	60	50	50	9	20	9	20	17	20
	Pigs per image	M	31.1	20.6	21.9	21.5	22.8	69.2	19.8	33.2	46.6	7.6	6.7
		SD	6.2	3.8	2.6	2.3	2.2	8.0	5.6	7.0	11.1	4.3	3.7
		MIN	20.0	13.0	12.0	16.0	19.0	57.0	7.0	26.0	27.0	3.0	1.0
		Q1	25.8	17.0	21.0	20.0	21.0	61.0	15.3	29.0	39.3	3.0	4.0
		Q2	32.5	21.0	22.0	22.0	22.5	70.0	21.0	32.0	44.0	6.0	6.0
		Q3	34.3	24.0	24.0	22.8	24.0	75.0	23.3	33.0	55.3	12.0	8.3
		MAX	45.0	26.0	27.0	27.0	27.0	79.0	28.0	50.0	65.0	14.0	15.0

image with minimal width and minimal height, when surrounding a pig completely. The classification task was defined as classifying the identified bounding box area of an image in one of the pre-defined object classes. The classes were (i) lying pigs vs. (ii) not lying pigs. Additionally, a real-valued confidence for the detection task was determined for evaluating the performance of our automatic detection approach. Summarizing, the object detection problem requires determining bounding box, corresponding object class and confidence score for each pig in a video image. Our approach for automatic object detection is based on deep neural networks (LeCun et al., 2015), which are trained and evaluated on an image dataset.

2.3. Dataset of annotated pig images

Our pig detection study was conducted using video images from pig fattening and pig rearing. Table 1 provides an overview of our pig image dataset. The table is structured hierarchically with the hierarchical level displayed in the left most column and relevant properties are provided for each hierarchical level in the rows of the table. The hierarchical levels are (1) compartments and contained pens, which were (2) recorded by cameras, which (3) generated one-hour long video recordings that were used to (4) extract a random sample of 10 images per video recording.

2.3.1. Pig compartments

The pens for pig fattening were located in building B and C. The camera perspectives, lighting and equipment in these pens were similar. The C pens (area of approximately 22 m^2) were smaller than B pens (area of approximately 33 m^2). All fattening pens contained a drinking station, which was located on the border to the dunging area, and a

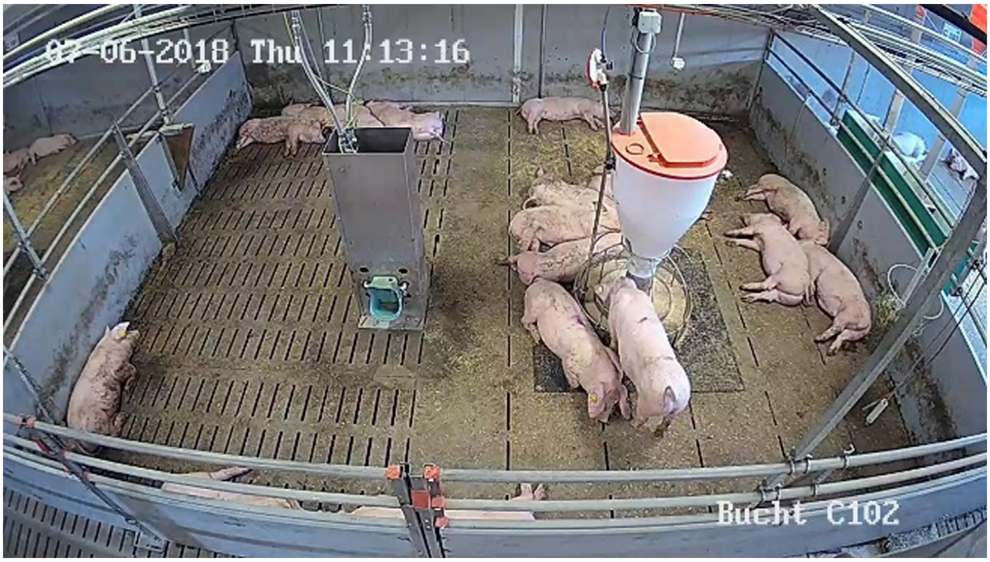

Fig. 1. A video image of compartment C1 with one completely visible fattening pen (C102).

mesh feeder in the lying area. Additionally, in all pens playing material such as hay, straw or different kinds of pellets were supplied to the pigs. For example, a swinging hay basket is located at the bottom right corner in Fig. 1. Fig. 1 depicts one of the C pens.

For three of the pens in compartment B1 one unique camera was inclined for recording more than three-quarters of the area of each pen. Additionally, one camera was located between the pens B101 and B102 (see Fig. 2) and one camera was located between the pens B103 and B104. The cameras between the pens covered about half of the area of each of the two pens.

The pens for pig rearing were located in building A and D. The camera perspectives, lighting and equipment in the buildings A and D were different in almost every aspect. The pen A1003 (area of approximately 15 m^2) was smaller than the pens in compartment D6 (area of approximately 23 m^2) and the shape was also different. The image dataset for pig rearing covers 2 pens in compartment D6 with similar lighting and equipment. These 2 pens were recorded by 3 cameras from 2 opposite inclined perspectives (i.e., from the front and the back) and 2 cameras in top-view. All these cameras did not record the complete pen. Both of the inclined camera perspectives recorded approximately half of the area of the pen. Fig. 3 provides an example of an inclined camera perspective of the rearing pens from compartment D6.

Additionally, we trained our pig detection approach using 17 images taken from two top-view cameras above the activity tower (a dispenser for playing material). The two cameras were installed above the two activity towers in the pens D601 and D602 (see Fig. 4). There were no other top view camera perspectives in the training set. In these video images about one quarter of the area of the pen was visible.

Finally, rearing pen A1003 was recorded by a different camera with a lower image resolution than the rest of the dataset (see Table 1). Fig. 5 provides an image of pen A1003 with the activity tower from a top-view perspective. The camera was located directly above it and recorded the pigs from a vertical perspective. Note that the posture of pigs is challenging to be (automatically) determined from the camera angle used.

2.3.2. Cameras

The cameras recorded in settings with a mixture of artificial light and day light. The HIKVision DS-2CD2125FWD-I IP Dome camera 2MP Full HD Outdoor 2.8 mm (HIKVision, Hangzhou, China) was used to

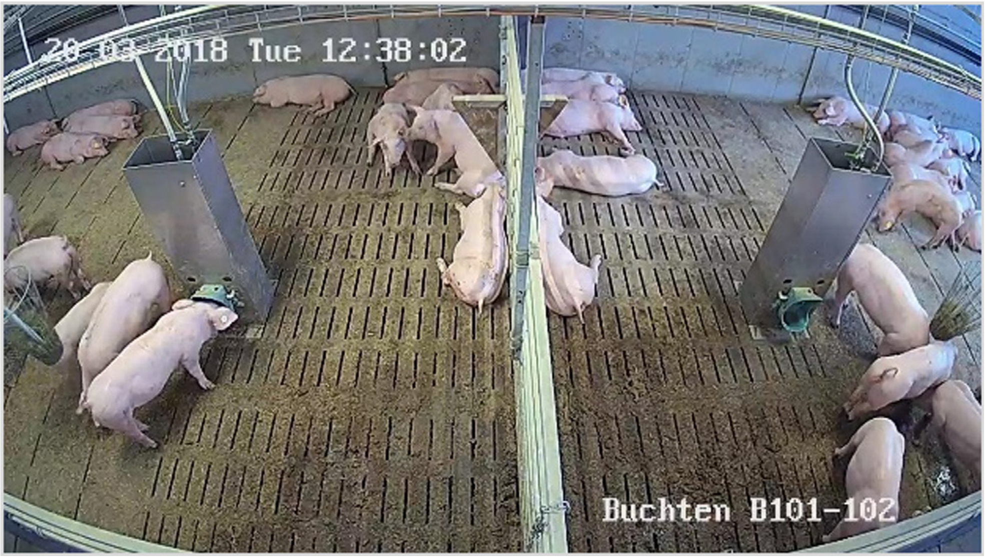

Fig. 2. A video image of compartment B1 with two visible fattening pens (i.e., B101, B102). The cameras between the pens recorded the slatted dunging areas.

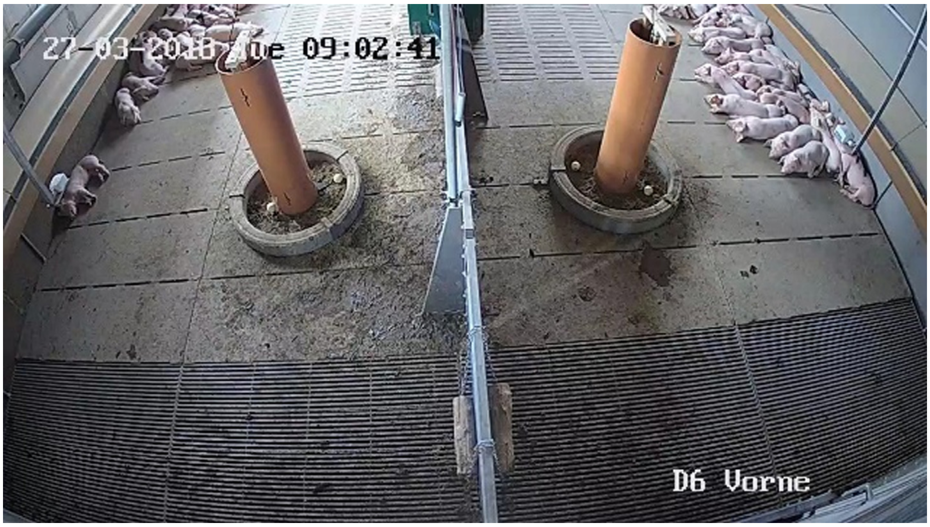

Fig. 3. An image taken by the front camera of compartment D6 with two visible rearing pens (i.e., D601, D602). Note the daylight that is visible in the left-top side of the image.

record videos in 1280 × 720 resolution. The recordings of pen A1003 were an exception because they have a resolution of 640 × 480.

The cameras were usually installed in a high angle shot, i.e., cameras were installed higher than the pigs and allow to look down on the pigs. Note that a high angle shot is not a completely vertically aligned view. A high angle shot is most convenient for human annotation in experiments, because it allows to cover a large area of pens without special equipment (e.g., fisheye lens) and the posture of pigs is clearly visible. A vertically aligned camera perspective, though, was used in case exact distances were required to be determined in the experiments. The videos were recorded during different stages of pig fattening and rearing in conventional pens. The cameras were installed such that they overlooked several pens. In some cases, the field of view covered only specific areas that were most relevant concerning third party research objectives.

2.3.3. Video image sample

To construct an image dataset, a random sample of 10 images per one-hour long video was drawn for each camera. 10 images were chosen as a balance between data management efforts and variability of the data. Our dataset consists of 31 videos, ensuring a high level of different pen settings. Because of high human effort in annotating images from those videos and the repetitive nature of the images from one video, we took a random sample of only 10 images per video.

Video recordings were split in two sets as noted in Table 1: (1) a set of images for training deep learning models for pig detection, and (2) a set of images for evaluating the models out of sample (i.e., the test set). To measure the out of sample performance of our object detection models, the holdout method was used, i.e., the test set was used to apply a trained model once for reliably estimating its application performance (Kohavi, 1995). The trained deep learning model was applied to the test set without further human interaction and without further adaptation of the model. We did not apply model selection and therefore did not require a validation set.

The dataset was designed to study the performance of our deep learning models given (1) similar training and test set (Section 3.1) and (2) a test set that is dissimilar to the training set to indicate the possibility to generalize for out of the box applications of other researchers (Section 3.2). First, to evaluate the performance given a large training set with similar images we used the videos recorded in the B and C building for pig fattening (see Table 1). The dataset for pig fattening contains 210 images of 16 different cameras that recorded 15 unique

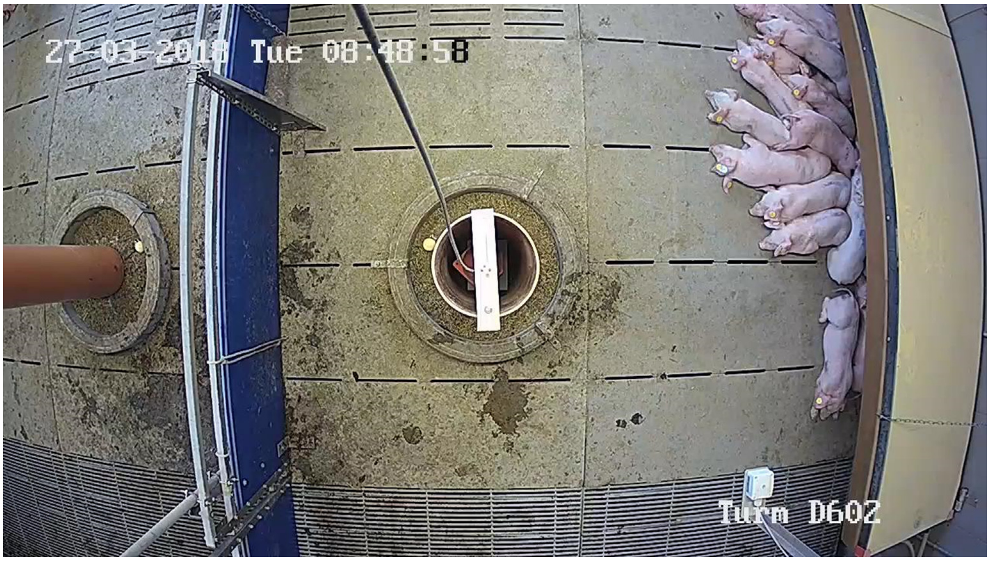

Fig. 4. A video image of the top-view of compartment D6 with two visible rearing pens (i.e., D601, D602).

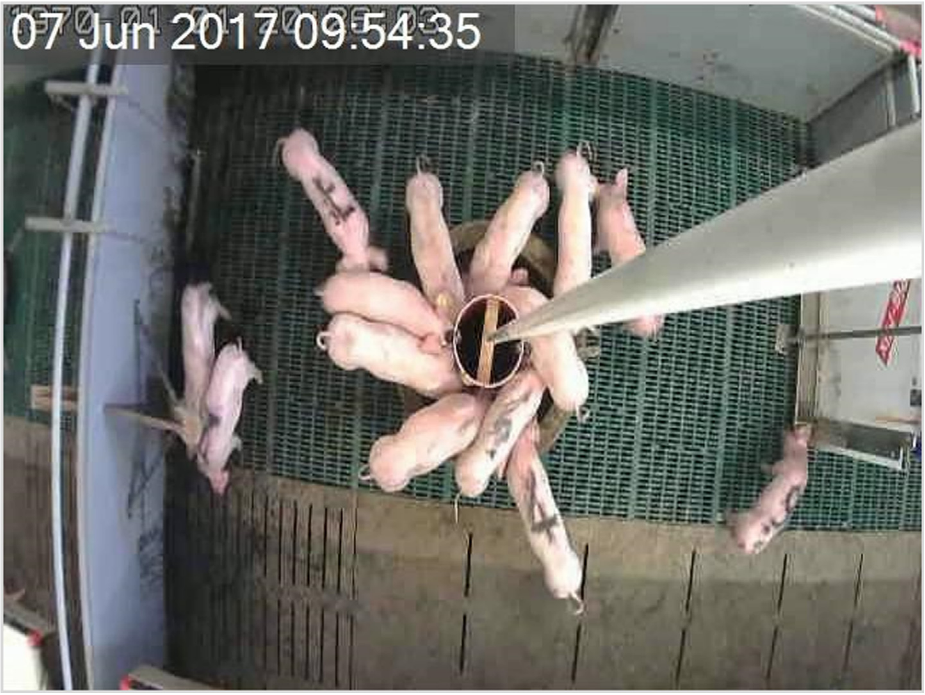

Fig. 5. An image of a recording of the top-view of rearing pen A1003.

pens. This dataset for pig fattening contains about two-thirds of all images used in our study. The camera perspectives and lighting conditions were similar in all cases. To evaluate our algorithms with unseen cameras the training set contained 0 images and the test set 50 images of 5 pens of the C2 compartment. Second, to evaluate our approach regarding unknown pen types and camera perspective the dataset for rearing pens was used. The dataset for pig rearing consists of 95 images, covering 3 different pens, recorded by 5 cameras. The camera and pen setting in A1003 were not in the training set. Therefore, the camera in A1003 (Fig. 5) recorded an unseen pen, camera type and image resolution perspectives (see Table 1). The top view camera perspective in A1003 was only present in 17 training images of a top view camera in pen D6 (Fig. 4). Hence, the performance on pen A1003 represents a scenario where our deep learning model is applied to a new pen without any training under vastly different conditions. We used the front and back cameras in D6 to study if 9 images for each camera is sufficient for training a deep learning model for dissimilar camera perspectives (see Table 1). Note that the shelter in D6 visible above the lying area at the left and right wall in the video recordings of the training set (see Figs. 3 and 4) was modified before the videos of the test set were recorded. In the test set a larger shelter was applied (see Figs. 11 and 12). This larger shelter was hiding a major part of the lying area in the test images of compartment D6.

The dataset was recorded at following times. The training set was recorded in B1 and C1 at 2018-03-20 between 11:05 and 12:50 and all pens in D6 at 2018-03-27 between 08:13 and 09:14. The test set was recorded in C1 and C2 at 2018-06-07 and 2018-06-08 between 11:00 and 12:09, in D6 2018-06-05 and 2018-06-15 between 8:00 and 12:57 and in pen A1003 at 2017-06-06 between 14:17 and 15:17 and 2017-06-07 between 09:18 and 10:18. The test set was recorded after the training set with the exception of pen A1003, which had no similar pen in the dataset. This kind of sampling prevents distorting the evaluation by images that are too similar, because they were recorded at the same day at a similar time. Please note that video recordings during times with high pig activity were selected. This is necessary, because pigs are lying most of the day, which would have further increased the imbalance of the dataset concerning the distribution of the classes lying and not lying. Training on such an imbalanced dataset biases the classifier towards the majority class ("pig lying") and an overly optimistic test performance would be achieved, because it is advantageous for the classifier to classify the more common class.

The images were assigned to the training set and test set by sampling videos as follows. The overall image dataset (including the training and test part) is a randomly stratified sample of the 31 video recordings with 10 images for each video recording. Of the 310 images in total, 5 had to be excluded due to privacy concerns, i.e., humans appeared in the images and could have been identified. All excluded images were in the training set. The final training set comprises 145 images and the test set comprises 160 images.

The descriptive statistics of the pigs per image in Table 1 were based on the annotations of the training and test set, i.e. the numbers describe the annotated pigs per image. Note that number of visible pigs in some of the video images were lower or higher than the number of housed pigs in the pens, because of (1) the number of visible pens in the camera (see Fig. 4), (2) the area of the pen that was visible in the video image (see Fig. 2), (3) pigs that were hidden behind objects (e.g., other pigs, the feeder, the drinker or pen barriers) and (4) neighboring pens that were visible in the video image, but did not count as visible pens because they were only visible in corners of the video image (see the left and right pens visible in Fig. 1). The annotation procedure is further described in the subsequent section.

2.3.4. Dataset annotations

Pigs in each of the 305 images were manually annotated. To draw the bounding boxes, we followed the guidelines of the reference challenge Pascal Voc 2010[1] (Everingham et al., 2010): A pig was considered truncated if the bounding box did not surround a pig fully in case (1) the pig extended outside an image, or (2) the pig was positioned behind a feeder or other object. If a pig was truncated, the bounding box was required to stop at the least visible part of the pig. A pig was considered occluded if parts of the pig were hidden. If a pig was occluded, the complete pig had to be surrounded by one bounding box because the edges of the pig were visible. These rules were closely followed

[1] http://host.robots.ox.ac.uk/pascal/VOC/voc2010/guidelines.html.

throughout the annotation process and were well known to each human annotator. We modified the Pascal Voc 2010 guideline for our annotation task in the following ways: (1) we did not mark any images as bad, because all video recordings had good image quality, and (2) we did not annotate occluded or truncated flags for the pigs and instead marked occluded and truncated pigs as regular annotations because our experiments did not require occluded or truncated flags. Furthermore, the annotations of pigs in neighboring pens were treated as regular annotations, which slightly increased the difficulty of our detection problem for the deep learning system, because these pigs were smaller and also truncated.

The training set was annotated by the corresponding author of the study. The images of the test set were annotated by two student assistants. Following the procedure of Everingham et al. (2010), no student annotated any images that were already annotated by the other student, i.e., each student annotated 5 images of each of the one-hour long videos. The annotations were checked by the corresponding author of the study to ensure thorough annotation as described by Everingham et al. (2010).

The tool "sloth"[2] was used for annotating pigs in images. The tool allows to select a pre-defined class and manually draw all bounding-boxes of an object in an image that represents this class. These bounding-boxes served as target variables for the deep learning algorithms. The manual annotations consist of a bounding box and an object class. We differentiated between lying, i.e., a pig was in a resting position and showed no muscular activity that was sufficient to detach its body from the surface, and not lying, i.e., any other activity. In total, 7277 bounding boxes were annotated, in which 5077 pigs were lying and 2200 pigs were not lying. In the training set, the number of annotations per classes were 849 pigs not lying and 2755 pigs were lying and in the test set 1,351 pigs were annotated not lying and 2322 pigs were annotated lying.

2.4. Deep learning system for automatic pig detection

We followed the current state of the art for object detection, i.e., creating a data-driven model using deep learning (LeCun et al., 2015). The term deep learning system refers to a feedforward neural network with a deep hierarchical structure (Schmidhuber, 2015). The images were provided to the model via an input layer. An output layer returned the detected bounding boxes with object class and confidence score. Between the input and output layers, hidden layers were used. Each layer consisted of nodes that were connected to the previous layers' nodes via weighted edges. The training phase used the pig annotations in the training set described in Section 2.3. During the training phase the weight parameters were iteratively optimized such that the model detected annotated pigs with minimal error (Schmidhuber, 2015). However, the training error is no reliable estimator of the out of sample performance of the model due to possible overfitting (Goodfellow et al., 2016). Overfitting occurs if the deep learning algorithm fits parameters to noise in the training set (Poggio et al., 2017). That is, the model may represent the training set well but may fail on prediction tasks on yet unseen data. To reduce overfitting, regularization techniques were applied, e.g. dropout (Srivastava et al., 2014). Additionally, the generalizability of the model was improved by convolution layers, i.e., a specific type of network layer (Chollet, 2017). An unbiased performance estimate regarding overfitting was achieved by out of sample evaluation on the test set as described in Section 2.4.

We used the following state-of-the-art architecture for our deep neural network. The Faster Region-based Convolutional Neural Network (Faster R-CNN) was used for object detection (Ren et al., 2015). Faster R-CNN achieved the highest mAPs in the Microsoft COCO: Common Objects in Context (MS COCO) reference challenge (Lin et al., 2014) given equal base networks, if computational resources were not restricted (Huang et al., 2016). The Faster R-CNN object detection pipeline defined the overall structure of the deep learning model. The Faster R-CNN is composed of three network modules that are stacked in the following order: (1) a base network, (2) a Region Proposal Network (RPN), and (3) an object classification network (Ren et al., 2015), providing detected and classified objects.

The base network was used to extract features from raw image data (Huang et al., 2016) and was derived from the first layers of the state-of-the-art image classification network called Neural Architecture Search (NAS) framework (Zoph et al., 2017). The RPN then proposes image regions that are likely to contain an object and the object classifier calculates the bounding boxes of the object and provides a soft maximum probability score for each possible class (Ren et al., 2015). The RPN and object classifier of Faster R-CNN were used (Ren et al., 2015). A visual overview of the Faster R-CNN object detection approach is provided by Zheng et al. (2018).

NAS represents the current state-of-the-art base network for object detection according to the reference challenge Microsoft COCO, in which the mAP of 0.431 was achieved using a Faster R-CNN architecture (Zoph et al., 2017). For example, Faster R-CNN (Ren et al., 2015) with NAS (Zoph et al., 2017) achieved higher mAP values than ResNet-101 (He et al., 2015), RetinaNet (Lin et al., 2017), R-FCN (Dai et al., 2016), YOLO-v3 (Redmon and Farhadi, 2018) and SSD (Liu et al., 2016) on the MS COCO reference challenge.

The following software and hardware were used during our experiments. We used tensorflow[3] v1.8 as open source deep learning software framework initially developed by Google Brain Team (Abadi et al., 2016) and the tensorflow object detection API[4] for the object detection task (Huang et al., 2016). We used the Nvidia CUDA v9.0.176 and Nvidia CUDNN v7.0 to run tensorflow on a Nvidia GeForce GTX 1080 Ti (MSI, Taiwan) graphics card.

For the hyperparameters of the deep learning model, the default configurations of the tensorflow object detection API were used, leaving random search-based optimizations to future work (Bergstra and Bengio, 2012). The weights of our deep learning model were pre-initialized by training on the MS COCO dataset to increase model performance (Lin et al., 2014). This was possible due to transferable characteristics in object detection tasks (Chollet, 2017). The learning rate was held constant at 0.0003 and the model was trained for 200,000 iterations. We applied a random horizontal flip for data argumentation, i.e., with a chance of 50% the ground truth image was flipped horizontally so that the right side of the ground truth image was visible on the left side of the training example and all annotated bounding boxes were flipped using the same approach.

We trained three deep learning models (DLM) for our experiments. The first model (DLM-1) was trained only on the training set to estimate pig position and posture. DLM-1 was used to calculate the mAP values for the posture detection. The second model (DLM-2) used only training data and was trained to estimate pig positions only, i.e., the training set contained only the class "pig". We used DLM-2 to calculate the position AP value in our experiments. The third system (DLM-3) was trained using the training and test data and was used for comparison to previous work (Sections 2.6 and 3.3) and exemplary application of the detection functionality (Section 3.3).

2.5. Evaluation procedures and metrics

In our study we used the following confusion matrix terminology for evaluating the deep learning models for pig detection (Manning and Schütze, 1999):

[2] https://github.com/cvhciKIT/sloth.

[3] https://www.tensorflow.org/.

[4] https://github.com/tensorflow/models/tree/master/research/object_detection

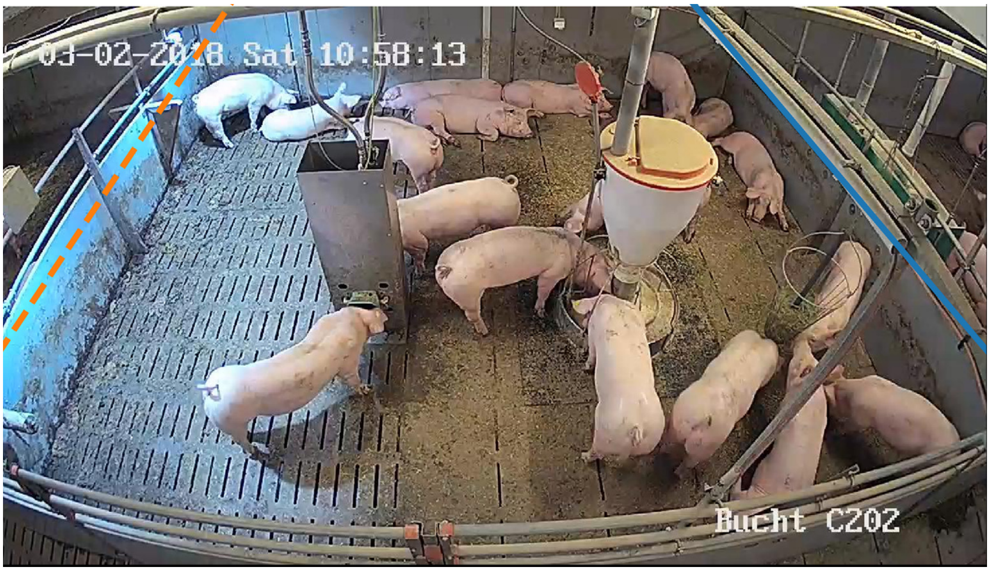

Fig. 6. Pig detections above the blue and orange lines were removed for the comparison with previous work and exemplary analysis of lying behavior of fattening pigs in pen C202. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

- *True positive (tp):* An object class is detected and the image contains this object class at this position.
- *False positive (fp):* An object class is detected while this object class is not at this position of the image.
- *False negative (fn):* An object class is at a certain position and it was not detected by the model.
- *True negative (tn):* No object class is at a certain position and the model did not detect an object class.

For the object detection task an additional parameter is required to calculate the variables above because an annotation and the predicted shape of the bounding box will never match perfectly. Therefore, it is necessary to define a minimum criterion, defining the accepted difference between ground truth and a detected bounding box. This parameter is called intersection over union (IOU) and it determines the required relative overlap α of the shape of the bounding boxes B_p and ground truth B_{gt} as defined by Everingham et al. (2015):

$$\alpha = \frac{area(B_p \cap B_{gt})}{area(B_p \cup B_{gt})}$$

The default value of this parameter is 0.5 (Everingham and Winn, 2010), which we used in our study.

Using the confusion matrix terminology and IOU, the following metrics were calculated (Manning and Schütze, 1999):

$$Recall = \frac{tp}{tp + fn}$$

$$Precision = \frac{tp}{tp + fp}$$

The average precision (AP) is an overall measure for the performance of an object detector concerning a specific class of an object detection task. The AP is calculated as follows: (1) ordering all pig detections based on their confidence score, (2) matching detections with highest confidence score to ground truth starting from highest confidence until a recall r higher than recall level $r$is reached and (3) calculating precision values based on each recall level r, and (4) interpolating the precision p_{interp} by the maximum precision that can be obtained for a recall level r as defined by Everingham et al. (2010):

$$p_{interp}(r) = \max_{r:r\geq r} p(r)$$

Eleven recall levels $r \in \{0, 0.1; \cdots, 1\}$ with equal distance were used (Everingham et al., 2010).

Finally, AP is the arithmetic mean of the precision p_{interp} at different recall levels according to Everingham et al. (2010):

$$AP = \frac{1}{11} \sum_{r \in \{0, 0.1; \ldots, 1\}} p_{interp}(r)$$

The mean average precision (mAP) is the mean of the AP values for each object class (Karpathy et al., 2014).

We implemented an additional evaluation procedure using 11 h of video data similar to Nasirahmadi et al. (2015). Nasirahmadi et al. (2015) studied group lying patterns in video recordings of top view cameras using a ellipse-based segmentation algorithm. Therefore, the objective of Nasirahmadi et al. (2015) is similar to our work. In the accompanying evaluation procedure, the number of housed pigs in the pen is compared with the detected number of pigs in a pen per image. This evaluation method is possible without a manually labeled dataset and can be applied to many hours of video data. Thus, we first recorded 11 h of continuous video data for pen C202. Pen C202 was used because the video recordings were of high quality, achieved high performance for posture detection and were similar to the other pens for pig fattening in our dataset. Thus, this example is well suited to demonstrate the possible application of the deep learning system. Second, we extracted one image every 10 s from each video, resulting in 3958 images. Third, we processed each of these images using DLM-3 and identified the position and posture of the pigs in each image. Because the model detects all pigs in an image, we removed all pig detections that were not part of pen C202. Concretely, we removed every detection where the lower-left edge of the bounding box was above the blue solid line on the right of Fig. 6. We also removed detections where the lower-right edge was above the orange dashed line. Then, we compared the number of detected pigs per image $\hat{y}_t$ at time t = 1, …, n to the number of actual pigs $y = 18$ that were housed in pen C202. The error was given by $e_t = y - \hat{y}_t$ and percentage error was defined by $p_t = 100e_t/y$ (Hyndman and Koehler, 2006). Finally, we reported the following metrics (Hyndman and Koehler, 2006).

Root Mean Squared Error (RMSE):

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(e_t)^2}$$

Mean Absolut Error (MAE):

$$MAE = \frac{1}{n}\sum_{t=1}^{n} |e_t|$$

Mean Absolut Percentage Error (MAPE):

$$MAPE = \frac{1}{n}\sum_{t=1}^{n} |p_t|$$

Accuracy:

$$Accuracy = 100 - MAPE$$

2.6. Exemplary application of detection functionality

We also studied pigs' lying behavior using the position and posture extracted from the eleven hours of continuous video data for pen C202. First, we used the detected position and posture data to generate an activity map of pigs, i.e., a spatial visualization of the areas where pigs most often were either lying or not-lying. For the activity map, we plotted a transparent colored rectangle for each detected pig in the shape of the bounding box. The rectangle was drawn in green if the detected class of a pig was lying and was drawn in blue if the detected class was not lying. For the background we used an image of the video recording and reduced its brightness to 40% of the original value to improve the visibility of the colors of the behavior analysis. The background image allows to see the pen layout and to interpret position and posture of the pigs with respect to the pen.

3. Results

3.1. Performance with high amount of training data for the case of fattening pigs

This section reports the results of the deep learning systems for pens with high amount of training data for the case of fattening pigs (see Table 1). Table 2 provides the AP and mAP values of the deep learning system for position and posture detection and the AP value for only detecting the pig position without detecting the posture (no classification). These metrics are calculated of images that were extracted of videos in the test set. Our system achieves a mAP of 80.9% for all pens in compartment C1. These pens were present in videos in our training set. A comparable mAP of 80.2% is achieved for the pens in compartment C2. The pens in compartment C2 were only present in our test set. Note that for both compartments the mAP values for position and posture detection and AP values for position detection are only different by less than 8%.

Fig. 7 provides a picture of the detection of only pig position of the deep learning system. The green bounding boxes mark pigs and only the class "pig" is classified using DLM-2. Above the green bounding box, the class name (left) and the confidence (right) of the deep learning system are depicted. Even pigs that are occluded by parts of an object could be detected. All pigs except one in the lower left part of the pen could be detected by the deep learning system.

Figs. 8 and 9 depict results of the deep learning system on images from the test set. The green bounding boxes mark pigs that are predicted to be lying and teal colored bounding boxes mark pigs that are predicted to be not lying. In Fig. 8 two pigs are visible that are truncated by the drinking station on the left side of the picture. One of these pigs was correctly detected by the deep learning system, but the other pig was not detected (i.e., a false negative).

Table 2
Evaluation results for test data with high amount of similar training data available for the respective fattening pen.

	Deep learning system	C1 (5 pens)	C2 (5 unseen pens)
AP only pig position	DLM-2	87.2%	87.4%
AP for the class "lying pig"	DLM-1	83.3%	78.5%
AP for the class "not lying pig"	DLM-1	78.6%	81.9%
mAP for pig position and posture	DLM-1	80.9%	80.2%

Fig. 9 depicts a pen from the compartment C2. Note the pig in the center, which is classified as lying and not lying in the same image. This occurs in rare cases if the detection algorithm is uncertain about the posture of the pig.

3.2. Results with little or no similar training data for the case of rearing piglets

In Table 3 we provide the corresponding results for pens with little or no training data. These pens are different from the majority of the training data concerning almost every aspect (see Table 1). Therefore, the detection of position and posture of rearing piglets in these video images is more challenging than position and posture detections of the previous fattening pigs. For the pen A1003 no training data was available and the class 'pig not lying' achieved an AP of 72.4%, but the AP of 'pig lying' is only 17.1%. Contrary, for the back camera of the compartment D6 the AP for 'pig lying' was 69.6% and 28.9% for 'pig not lying' was achieved. The most likely explanation for these opposing AP values were the different camera perspectives used, which may have had different effects on the classification network of DLM-1. The AP of the pig position without posture detection was 76.8% for A1003 and 73.6% for the back camera of compartment D6.

Fig. 10 depicts the performance of the pig position detection without posture detection by DLM-2. Fig. 10 demonstrates that the position of pigs that were close together was detected individually even for camera perspectives for which only little training data was available.

Fig. 11 shows position and posture detections for the pen A1003. Note that the confidence scores (visible in the image on the right side of the bounding box) of DLM-1 for the detections for this camera perspective were lower for some of the detections. This indicates that the algorithm was uncertain concerning these detections.

Fig. 12 shows the detections of DLM-1 for the pens in compartment D6 recorded by the camera in the back of the pen. Note the lying pig detections in the center of the image. For this camera perspective not lying piglets were often classified as lying if the legs of the pig were truncated by other piglets. It can be assumed that this specific posture was not part of the training data in the fattening pens in building C, because in the pens in building C the feeder of the pen was most often round and accessible from all sides (see Fig. 9) or it was shorter and located at the side of the pen, i.e. at a different angle (see Fig. 8).

3.3. Exemplary application using 11 hours of video data

First, we implemented the evaluation procedure for the number of detected pigs compared to housed pigs per image as described in Section 2.5 using the 11 h of video recordings of pen C202 for the deep learning system DLM-3. The RMSE was 1.62, the MAE was 1.25, the MAPE was 6.9% and the accuracy was 93.1%. The MAE implies that the number of detected pigs was on average 1.25 pigs higher or lower than the 18 pigs housed in the pen.

Our deep learning system DLM-3 was used to study the lying behavior of fattening pigs as depicted in the activity map in Fig. 13. The green colored rectangles marked the area where pigs were detected to be lying, while the blue colored rectangles marked pigs that are detected to be not lying. The detected behavior of the pigs in the video shows that pigs were lying (green) at the walls of the pen and outside of the dunging area. Not-lying pigs (blue) were observed in the area between the drinker and the feeder. Pigs were rarely observed in the dunging area and were mostly not lying in the dunging area (almost no

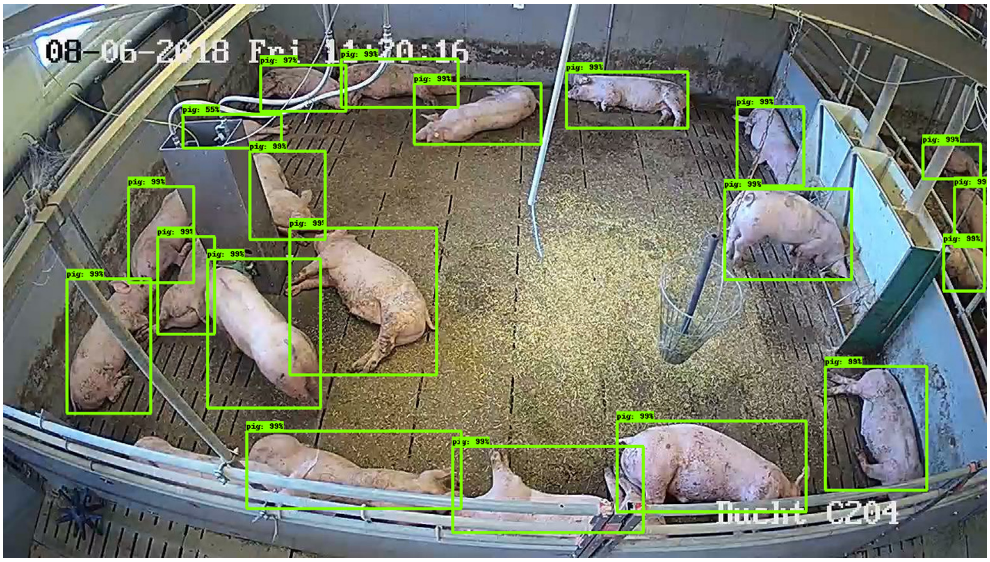

Fig. 7. Detections of pig position of the deep learning system on a test set image for fattening pen C204. Images of this pen were not in the training set.

green and only slightly blue).

Fig. 14 depicts a time series of activity during the 11 h of video data detected by the deep learning system DLM-3. These 11 h of video data were prepared as described in Section 2.6. In the next step we plotted a green time series for lying pigs and a blue time series for not lying pigs. The x-axis depicts the daytime of the analyzed video image and the y-axis shows the number of pigs.

Fig. 15 depicts a period were no pig was detected lying by DLM-3 in pen C202 and Fig. 16 shows a period where no pig was detected not lying by DLM-3 in pen C202. For Fig. 15 these detections were correct for all pigs in pen C202. For Fig. 16 one detection error is visible at the feeder of the pen.

4. Discussion

4.1. Findings

This work evaluates deep learning for detecting the position and posture of pigs in images from standard 2D cameras. Previous work achieved this in a top view perspective (Nasirahmadi et al., 2015). Our deep learning system contributes by detecting postures from real-world non-vertical camera perspectives using 21 cameras in 18 pens. We evaluated the predictions of the system regarding pigs' position and posture with the state-of-the-art mAP metric for object detection. The position and posture of pigs can be detected with 80% mAP for camera perspectives that are not entirely vertically directed and for which sufficient training data are provided. A similar work detected the lying behavior of one lactating sow from one top view 3D camera for five posture types using deep learning with 87.1% mAP (Zheng et al., 2018). However, this mAP value is not comparable to our work because of different experimental parameters. Position and posture detection for experiments with little training data and a high angle shot perspective achieved 49.2% mAP and 58.8% mAP. Position and posture detection with no training data and a difficult top view camera perspective achieved 44.8% mAP. Finally, for the detection of the position of a pig without posture, an AP of over 87% with high amount of training data and over 67% for little or no training data was found. These results indicate that the position of a pig without posture can be detected well for different unseen camera perspectives and pen layouts.

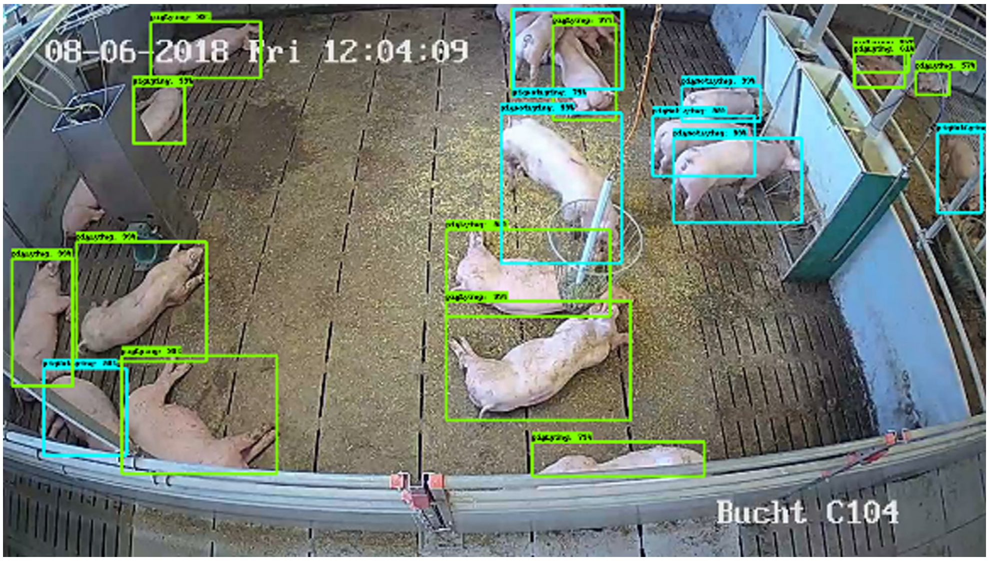

Fig. 8. Detections of the deep learning system on a test set image for fattening pen C104. Images of this pen were in the training set.

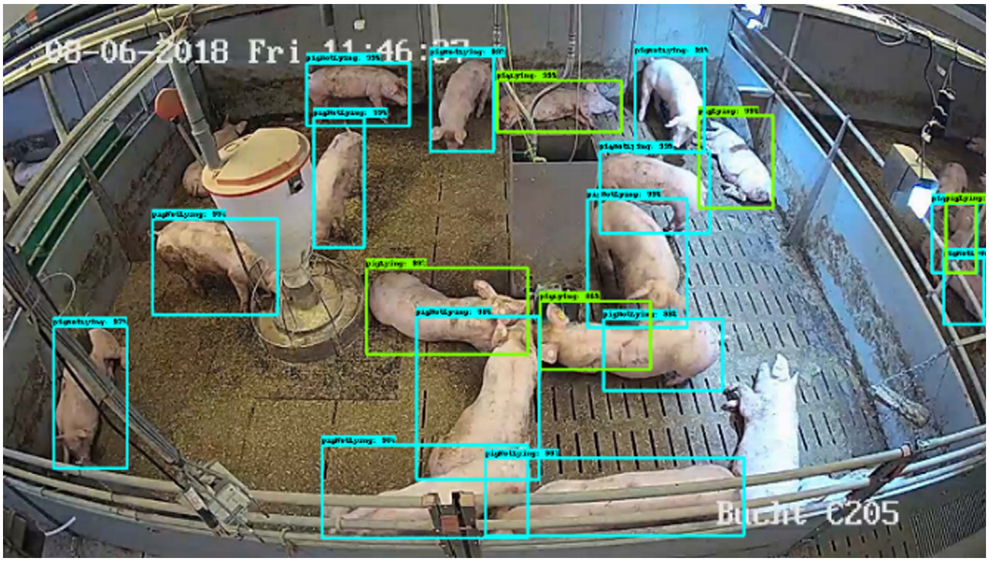

Fig. 9. Detections of the deep learning system on a test set image for fattening pen C205. No images of this pen were in the training set.

Table 3
Evaluation results for test data with little or no training data available for the respective rearing pen.

	Deep learning system	D6 back camera (18 similar training images)	D6 front camera (18 similar training images)	A1003 (no similar training images)
AP only pig position	DLM-2	73.6%	67.7%	76.8%
AP for the class “lying pig”	DLM-1	69.6%	53.4%	17.1%
AP for the class “not lying pig”	DLM-1	28.9%	62.5%	72.4%
mAP for pig position and posture	DLM-1	49.2%	58.8%	44.8%

4.2. Limitations

Our research has several limitations. First, in future investigations the performance of different deep learning systems should be compared. In this work we used the state-of-the-art Faster R-CNN object detection pipeline and the state-of-the-art NAS base network. However, model selection concerning different base networks and object detection pipelines should be studied using our publicly available dataset, because the performance of a machine learning algorithm varies for different detection problems (Wolpert, 1996). Additionally, each deep learning network comes with various hyperparameters (e.g., learning rate and number of training steps) for which we used the default configuration. The reason for not applying model selection including hyperparameter optimization is that model selection requires an additional validation set, i.e., a dataset specifically for model selection. This is necessary because measuring the performance of several models and choosing the model with highest performance will result in overfitting of the chosen model to the validation set. Therefore, if model selection

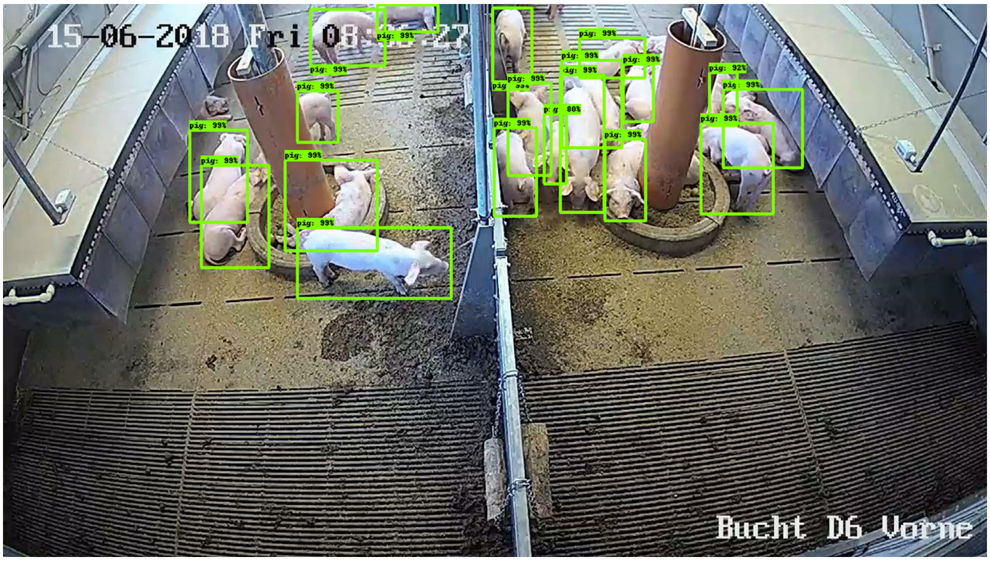

Fig. 10. Detections of pig position of the deep learning system on a test set image for rearing pens in the compartment D6 (back camera). The training set contained 18 images of this pen type and 9 from this camera.

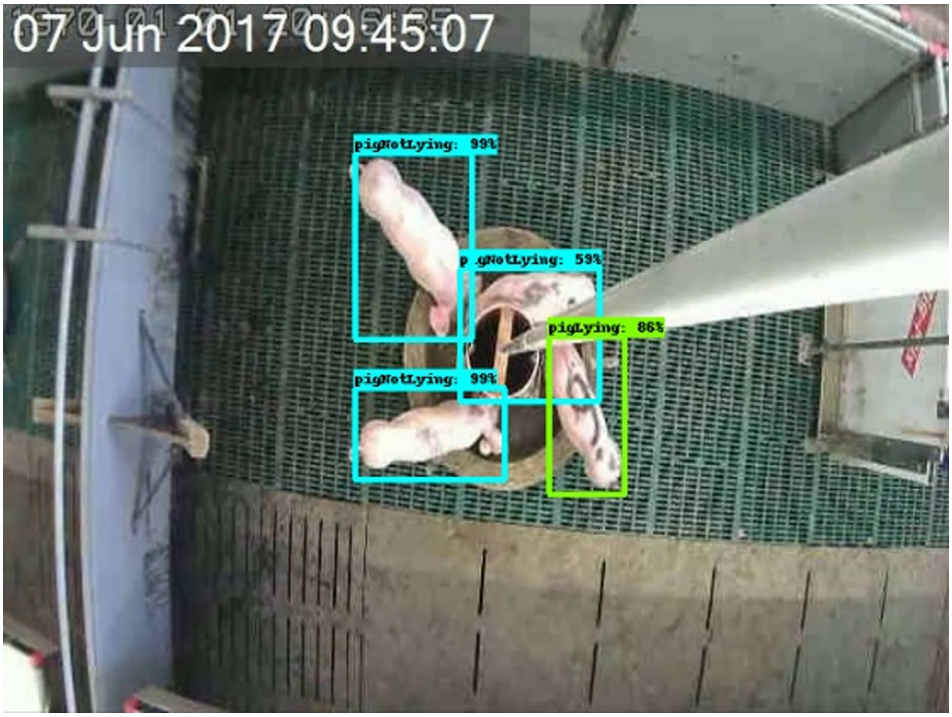

Fig. 11. Detections of the deep learning system on a test set image for rearing pen A1003. No images of this pen type or camera were in the training set. The training set contained 17 images of a top view perspective from another compartment.

would have been applied using the test set as the validation set, the experiments in Sections 3.1 and 3.2 would not have been possible on an unbiased estimate of the actual performance of the model (the test set would have been already used as the validation set). Therefore, given the provided dataset the most valuable contribution was achieved by deliberately excluding model selection from these experiments. However, by making our dataset publicly available, we allow for model selection experiments to further improve the performance. Researchers can use our data as a training set and may annotate a validation and test set based on images of video recordings from unseen pens.

Second, while our AP for position detection on the test set with little or no training data for rearing pens was satisfactory, the mAP for posture detection was lower than expected (Section 3.2). Degrading performance with no similar training data is in accordance with previous work that used deep learning for detecting the feeding behavior of pigs (Yang et al., 2018). Transferring the trained detection algorithm on different piggeries was not successful (Yang et al., 2018). It is worth noting that in their work another dataset, base network and hyperparameter configuration was used, i.e., Faster R-CNN with the base network ZF-Net and a (different) learning rate of 0.001. A possible explanation for the degraded performance on dissimilar test images might be that the training data used in our study and the study of Yang

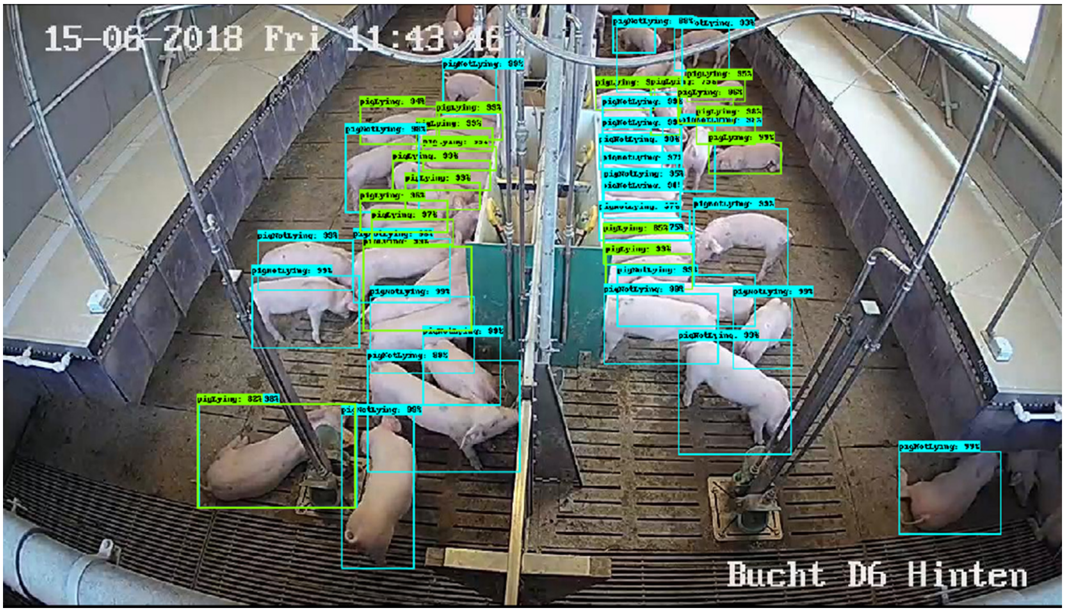

Fig. 12. Detections of the deep learning system on a test set image for a pen in the rearing compartment D6. The training set contained 18 images of this pen type and camera perspective.

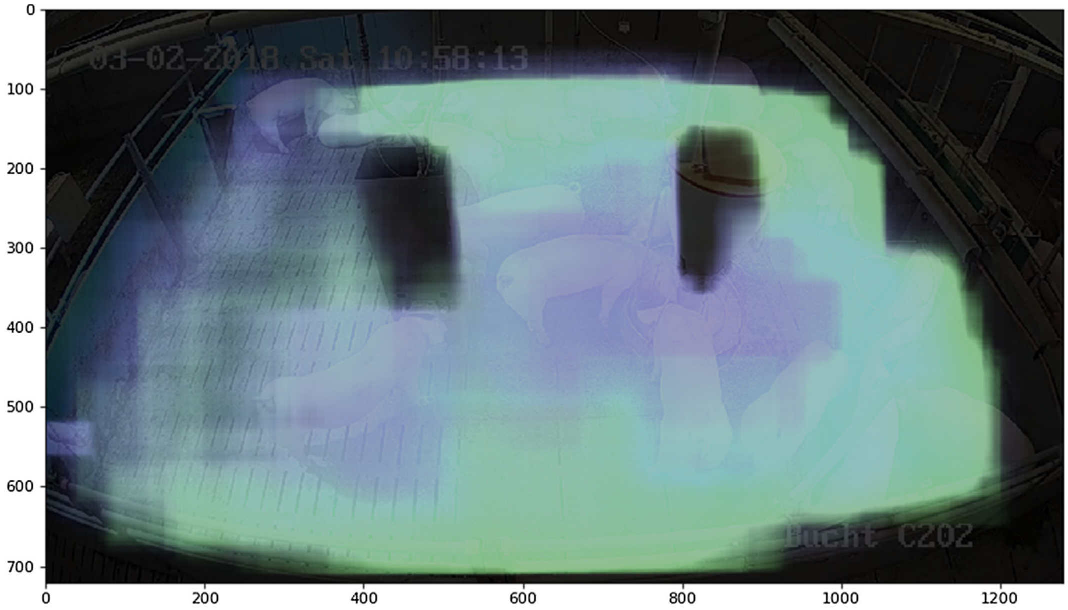

Fig. 13. Activity map of position and posture of pigs during 11 h of video recording for the fattening pen C202. Green color depicts lying pig locations and blue depicts not lying pigs. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

et al. (2018) did not exhibit sufficient variety, e.g., concerning the background, camera type and camera perspective. In our work, the variety was increased by including 15 cameras and 12 pens in the training set. However, most reference datasets for object detection (e.g., Microsoft Common Objects in Context) have higher variety in their images, because their data stems from internet image databases (e.g., flickr) and each image is unique regarding every aspect (Lin et al., 2014). Another explanation might be that our model has not been specifically selected to achieve high performance for pig posture detection. However, this explanation seems unlikely because we achieve high mAP values (80.2%) given enough similar training data even for the unseen pens in the C2 compartment (Table 2) and high AP values for position detection for little and no training data (Table 3). Please also note that the similar mAP between seen C1 (80.9%) and unseen C2 (80.2%) pens indicates that among similar conditions with enough training data the deep learning models do not indicate overfitting (Table 2). However, model selection or hyperparameter optimization on the available dataset might improve the performance on the pens in the validation set, but could also degrade the performance on unseen pens due to overfitting. Thus, fellow researchers who apply our deep learning system to new pens and camera perspective might not achieve our results on their data for posture detection. To overcome this problem, we recommend annotating about 50–100 images of an unseen pen, adding them to our training set, and fine-tuning the deep learning system.

Third, the percentage of not lying pigs was 23.6% in the training set and 36.8% in the test set respectively the difference is largely explained by the high number of not-lying pigs in rearing pens in the test set, which showed a not-lying percentage of 61.9% compared to 15.5% not-lying pigs in rearing pens in the training set. An explanation for the difference in lying percentage might be that the large lying area in the rearing pens visible in Figs. 3 and 4 at the left and right wall was hidden under the closed and larger shelter visible in the area under Figs. 11 and 12, which covered most of the lying pigs. Please note that the discussed difference in percentage of not-lying pigs in the rearing pens may also partly explain the decreased performance of the position and posture detection for rearing pens discussed in the second limitation.

Fourth, we chose video recordings during times with high pig activity in our dataset. Therefore, the provided mAP values are not an estimate of the performance during a complete day. However, a more balanced class distribution (during times with high pig activity) would be more challenging for detection tasks, because deep learning systems are biased towards detecting the majority class of the dataset. Therefore, mAP evaluated during a complete day might increase,

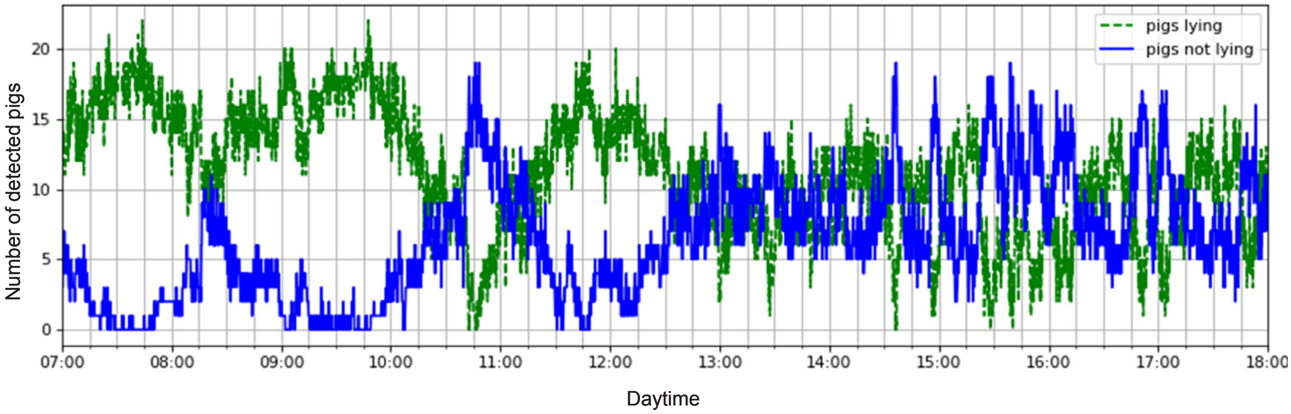

Fig. 14. Pig lying behavior within 11 h of video recording for the fattening pen C202. Green depicts lying pigs and blue depicts not lying pigs. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

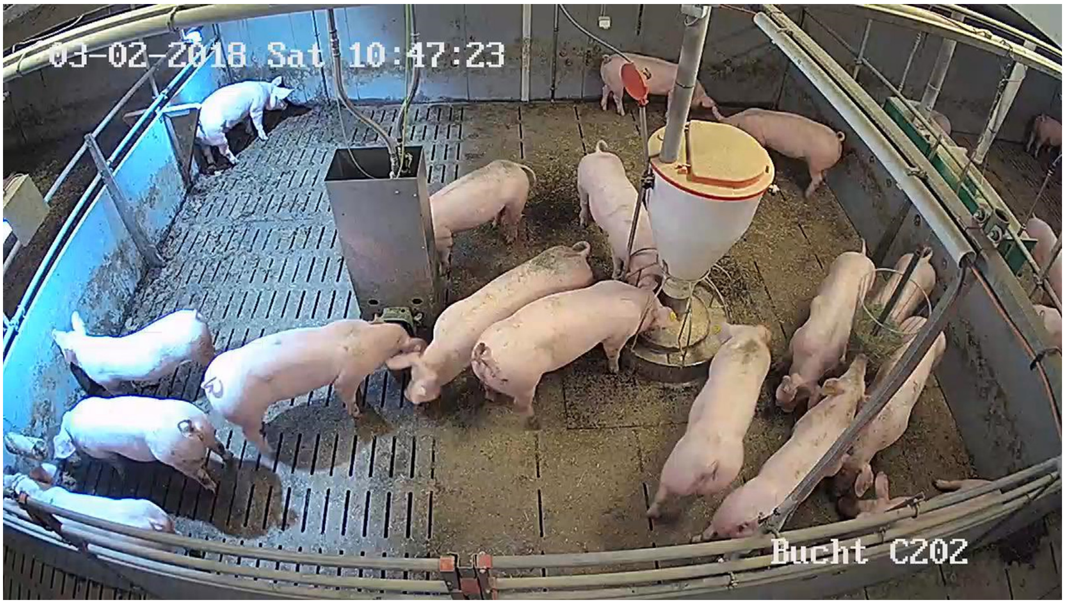

Fig. 15. Highest recorded pig activity during this day at 10:47 for the fattening pen C202. No pigs were detected lying at this time.

because the majority class (pig lying) is more likely to occur.

Fifth, our exemplary application of the detection functionality of the pig lying behavior only includes eleven hours of video data of one pen, and is thus not a reliable general estimation of pig behavior. However, our analysis demonstrates possible use of the deep learning system in behavioral studies as described in Section 4.3.

4.3. Implications

Our study has the following implications for research and practice. First, future work could optimize our deep learning models using model selection and identify how hyperparameters should be configured. For example, the learning rate at 0.0003 and the 200,000 iterations were the baseline configurations of the tensorflow object detection API. An inspection of the learning curves of our deep learning models indicates that the training error converges early on during the training phase. Therefore, the number of iterations might be reduced during the training phase to accelerate model selection. Furthermore, varying the learning rate might improve performance. A chance of 50% horizontal flip is chosen in the baseline configuration of the tensorflow object detection API, because it increases the variety in the training set without providing a downside in the pig position and posture detection task and is a default procedure for object detection (Girshick, 2015). Therefore, there is no advantage to select the chance for a horizontal flip experimentally. Furthermore, different base networks (e.g., ResNet-101, Inception-ResNet-v2) or object detection pipelines (e.g., SSD or R-FCN) could be applied.

Second, our experiments in Section 3.2 indicate that applying model selection or increasing the size of the training set without increasing the variety of the training set with the aim of increasing the mAP for posture detection is unlikely to provide generalizable deep learning models for out of the box applications (see second limitation). Therefore, the implication for future work is that increasing the variety of experimental conditions in the training images in almost every aspect will

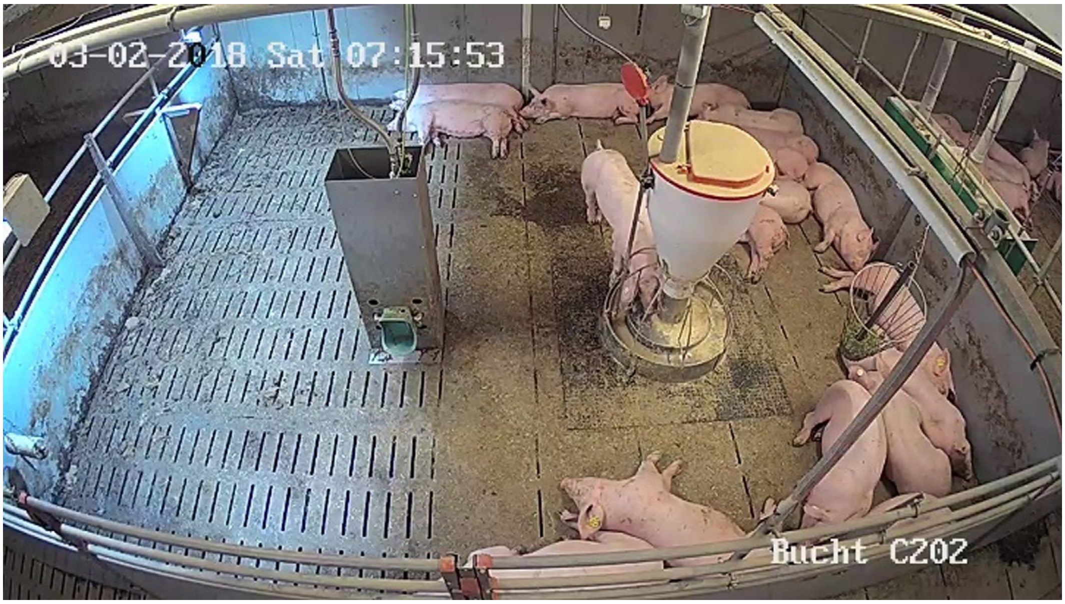

Fig. 16. No pigs were detected not lying by the deep learning system at 07:16 for the fattening pen C202.

likely be necessary for out of the box posture detection applications. This is difficult to achieve by individual groups of researchers, e.g. we used the majority of available cameras with different experimental conditions in our research facility to create our dataset. Therefore, providing publicly available datasets from a large number of piggeries might be necessary for out of the box posture detection. Hence, the short-term application for posture detection are research settings where the number of studied pens is exhaustive (see third implication). In this setting, our publicly provided dataset reduces the amount of training data, because it can be added to the training set of the studied pens.

Third, a further study with a focus on animal behavior is suggested. For this purpose, position and posture detection functionality can be utilized over long periods of time (e.g., months or years) in short intervals (e.g., every minute). The position detection functionality (i.e., the rectangular area defined by two vectors for each pig) can be used to monitor activity hotspots and the distance of pigs to an area of interest (e.g., drinker or feeder) throughout the day. This position detection functionality can be combined with the lying detection functionality for each pig, e.g. to count the number of lying pigs in an area or distance of lying pigs from hotspots. Furthermore, the detection time provides additional functionality and allows the analysis of time series data concerning animal behavior as demonstrated in Section 4.3. Therefore, this further study could analyze several months of video data targeting the prediction of tail biting, early-disease detection, correct use of the functional areas of the pen, and other animal welfare parameters. For example, it would be interesting to further analyze the time series in Section 3.3 and study if animal welfare threatening behavior can be detected via time series patterns. In Fig. 15 at 10:47 h we can observe new food being delivered to the feeding station, which increased the activity of the pigs. Note that in Fig. 14 the peak is gradually increasing and decreasing. Between 14:30 and 16:15 several peaks of activity are visible in the time series. Viewing the video between 14:30 and 16:15 shows that at that time aggressive behavior occurred. It might be that these short spikes in activity are some indicators of unwanted behavior. Second, the exact lying behavior classified by our deep learning system could help to design better pens. Activity maps as demonstrated in Fig. 13 can be used to better understand pig behavior in different types of pens.

Fourth, the position and posture detection could also be used for an automated animal welfare monitoring system for rearing and fattening pigs that could be used to detect aggression or disease of pigs. The advantage of image analysis is that animal behavior can be recorded and analyzed without individual and invasive sensors for each pig. Further work could study how to apply position and posture detection for this application. For this purpose, the deep learning system should be generalized to night recordings, additional camera perspectives and pen designs.

5. Conclusion

Our work contributes to the understanding of how to adopt deep learning for detecting the position and posture of pigs using 21 cameras in 18 pens under realistic conditions. Positions were detected with over 67.7% AP in all our experiments. Position and posture detection led to a mAP of 80.2% for unseen pens with enough similar training images. For detecting position and posture with only 0 or 18 similar training images, the mAP was lower, ranging between 44.8% and 58.8%. The latter position and posture detection results are not sufficient for most out-of-the-box posture detection applications. Our results indicate that increasing the variety of experimental conditions of the training images might be necessary to increase transferability of the detection models. Therefore, further research and open access datasets of different piggeries are needed to ensure the performance on new camera perspectives or pen layouts. One advantage of deep learning is that the applicability to new camera perspectives or pen layouts can iteratively be improved by adding annotations for new camera perspectives or pen layouts and fine-tuning the deep learning system to this training data. Specifically, in a research setting with an exhaustive set of similar pens the approach can be applied with limited effort for position and posture detection. Practitioners may apply position detection, which may work out-of-the-box as our limited experiments suggest. Furthermore, we propose exemplary applications that can support the pen design by visualizing the location and the active times during the day as a further contribution. Thus, our research helps to shorten the path for practitioners and researchers to design precise detection models in the context of improving animal welfare. The dataset used in our experiments is publicly available and can be retrieved from https://wi2.uni-hohenheim.de/analytics.

CRediT authorship contribution statement

Martin Riekert: Writing - original draft, Writing - review & editing, Conceptualization, Methodology, Investigation, Data curation, Software, Formal analysis, Validation. **Achim Klein:** Writing - original draft, Writing - review & editing, Supervision, Conceptualization, Methodology, Funding acquisition. **Felix Adrion:** Writing - original draft, Writing - review & editing, Conceptualization, Methodology. **Christa Hoffmann:** Writing - original draft, Writing - review & editing, Project administration, Funding acquisition, Resources. **Eva Gallmann:** Writing - review & editing, Supervision, Conceptualization, Funding acquisition, Data curation, Resources.

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

The authors would like to thank the staff at the Boxberg Teaching and Research Centre – Centre for pig rearing and pig breeding (LSZ Boxberg) for their support and advise during the experiments, especially Hansjörg Schrade, Barbara Keßler, Andrea Wild, Max Staiger, Karen Kauselmann and William Gordillo. We also thank the student assistants Tobias Zimpel and Endrit Tejeci for annotating the test set used in our studies. We thank the LabelFit project team, especially Svenja Opderbeck and Benedikt Glitz, who allowed us to use their video dataset. We thank Prof. Dr. Stefan Kirn and Dr. Marcus Müller (University of Hohenheim) for discussions and ideas. We thank Dr. Jörg Leukel (University of Hohenheim) for his comments on an earlier version of the manuscript. Finally, we thank the machine learning community for providing many free resources for our research.

Funding

This work was supported by the project "Landwirtschaft 4.0: Info-System", funded by the Ministry of Rural Development and Consumer Protection of Baden-Württemberg, Germany. This work was also supported by the project LabelFit by funds of the Federal Ministry of Food and Agriculture (BMEL) based on a decision of the Parliament of the Federal Republic of Germany via the Federal Office for Agriculture and Food (BLE) under the innovation support programme (2819200415).

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., et al., 2016. TensorFlow: large-scale machine learning on heterogeneous distributed systems. Retrieved from. ArXiv. http://arxiv.org/abs/1603.04467.

Adrion, F., Kapun, A., Eckert, F., Holland, E.-M., Staiger, M., Götz, S., Gallmann, E., 2018. Monitoring trough visits of growing-finishing pigs with UHF-RFID. Comput. Electron. Agric. 144, 144–153. https://doi.org/10.1016/j.compag.2017.11.036.

Ahrendt, P., Gregersen, T., Karstoft, H., 2011. Development of a real-time computer vision system for tracking loose-housed pigs. Comput. Electron. Agric. 76 (2), 169–174. https://doi.org/10.1016/j.compag.2011.01.011.

Bergstra, J., Bengio, Y., 2012. Random search for hyper-parameter optimization. J. Mach. Learn. Res. 13, 281–305.

Borchers, M.R., Chang, Y.M., Tsai, I.C., Wadsworth, B.A., Bewley, J.M., 2016. A validation of technologies monitoring dairy cow feeding, ruminating, and lying behaviors. J. Dairy Sci. 99 (9), 7458–7466. https://doi.org/10.3168/jds.2015-10843.

Brendle, J., Hoy, S., 2011. Investigation of distances covered by fattening pigs measured with VideoMotionTracker®. Appl. Anim. Behav. Sci. 132 (1–2), 27–32. https://doi.org/10.1016/j.applanim.2011.03.004.

Brünger, J., Traulsen, I., Koch, R., 2018. Model-based detection of pigs in images under sub-optimal conditions. Comput. Electron. Agric. 152, 59–63. https://doi.org/10.1016/j.compag.2018.06.043.

Chollet, F., 2017. Deep Learning with Python. Manning Publications.

Condotta, I.C.F.S., Brown-Brandl, T.M., Silva-Miranda, K.O., Stinn, J.P., 2018. Evaluation of a depth sensor for mass estimation of growing and finishing pigs. Biosyst. Eng. 173, 11–18. https://doi.org/10.1016/j.biosystemseng.2018.03.002.

D'Eath, R.B., Jack, M., Futro, A., Talbot, D., Zhu, Q., Barclay, D., Baxter, E.M., 2018. Automatic early warning of tail biting in pigs: 3D cameras can detect lowered tail posture before an outbreak. PLoS One 13 (4), 1–18. https://doi.org/10.1371/journal.pone.0194524.

Dai, J., Li, Y., He, K., Sun, J., 2016. R-FCN: object detection via region-based fully convolutional networks. Advances in Neural Information Processing Systems (NIPS).

Everingham, M., Eslami, S.M.A., Van Gool, L., Williams, C.K.I., Winn, J., Zisserman, A., 2015. The Pascal visual object classes challenge: a retrospective. Int. J. Comput. Vision 111 (1), 98–136. https://doi.org/10.1007/s11263-014-0733-5.

Everingham, M., Van Gool, L., Williams, C.K.I.I.I., Winn, J., Zisserman, A., Luc, et al., 2010. The PASCAL visual object classes (VOC) challenge. Int. J. Comput. Vision Manuscript 88 (2), 303–338. https://doi.org/10.1007/s11263-009-0275-4.

Everingham, M., Winn, J., 2010. The PASCAL Visual Object Classes Challenge 2010 (VOC2010) Development Kit (Vol. 2010). Leeds.

Girshick, R., 2015. Fast R-CNN. In: 2015 IEEE International Conference on Computer Vision (ICCV). IEEE, pp. 1440–1448. https://doi.org/10.1109/ICCV.2015.169.

Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep Learning. MIT Press.

Guo, Y., Zhu, W., Jiao, P., Ma, C., Yang, J., 2015. Multi-object extraction from topview group-housed pig images based on adaptive partitioning and multilevel thresholding segmentation. Biosyst. Eng. 135, 54–60. https://doi.org/10.1016/j.biosystemseng.2015.05.001.

Hammer, N., Adrion, F., Staiger, M., Holland, E., Gallmann, E., Jungbluth, T., 2016. Comparison of different ultra-high-frequency transponder ear tags for simultaneous detection of cattle and pigs. Livestock Sci. 187, 125–137. https://doi.org/10.1016/j.livsci.2016.03.007.

Hammer, N., Pfeifer, M., Staiger, M., Adrion, F., Gallmann, E., Jungbluth, T., 2017. Cost-benefit analysis of an UHF-RFID system for animal identification, simultaneous detection and hotspot monitoring of fattening pigs and dairy cows. Landtechnik 72, 130–155. https://doi.org/10.15150/lt.2017.3160.

He, K., Zhang, X., Ren, S., Sun, J., 2015. Deep residual learning for image recognition. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–17. https://doi.org/10.1007/s11042-017-4440-4.

Huang, J., Rathod, V., Sun, C., Zhu, M., Korattikara, A., Fathi, A., et al., 2016. Speed/accuracy trade-offs for modern convolutional object detectors. In: IEEE Comput. Vision Pattern Recognit.pp. 3296–3297. https://doi.org/10.1109/CVPR.2017.351.

Hyndman, R.J., Koehler, A.B., 2006. Another look at measures of forecast accuracy. Int. J. Forecast. 22 (4), 679–688. https://doi.org/10.1016/j.ijforecast.2006.03.001.

Kalbe, C., Zebunke, M., Lösel, D., Brendle, J., Hoy, S., Puppe, B., 2018. Voluntary locomotor activity promotes myogenic growth potential in domestic pigs. Sci. Rep. 8 (1), 1–11. https://doi.org/10.1038/s41598-018-20652-2.

Kamilaris, A., Prenafeta-Boldú, F.X., 2018. Deep learning in agriculture: a survey. Comput. Electron. Agric. 147 (1), 70–90. https://doi.org/10.1016/j.compag.2018.02.016.

Kapun, A., Adrion, F., Gallmann, E., 2018. Activity analysis to detect lameness in pigs with a UHF-RFID system. In: 10th International Livestock Environment Symposium (ILES X). American Society of Agricultural and Biological Engineers, St. Joseph, MI. https://doi.org/10.13031/iles.18-068.

Karpathy, A., Toderici, G., Shetty, S., Leung, T., Sukthankar, R., Fei-Fei, L., 2014. Large-scale video classification with convolutional neural networks. In: IEEE Conference on Computer Vision and Pattern Recognition. IEEE, pp. 1725–1732. https://doi.org/10.1109/CVPR.2014.223.

Kashiha, M.A., Bahr, C., Ott, S., Moons, C.P.H.H.H., Niewold, T.A., Tuyttens, F., Berckmans, D., 2014. Automatic monitoring of pig locomotion using image analysis. Livestock Sci. 159 (1), 141–148. https://doi.org/10.1016/j.livsci.2013.11.007.

Kashiha, M., Bahr, C., Ott, S., Moons, C.P.H.H., Niewold, T.A., Ödberg, F.O.O., Berckmans, D., 2013. Automatic identification of marked pigs in a pen using image pattern recognition. Comput. Electron. Agric. 93, 111–120. https://doi.org/10.1016/j.compag.2013.01.013.

Kim, J., Chung, Y.Y., Choi, Y., Sa, J., Kim, H.H., Chung, Y.Y., et al., 2017. Depth-based detection of standing-pigs in moving noise environments. Sensors 17 (12), 2757. https://doi.org/10.3390/s17122757.

Kohavi, R., 1995. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: International Joint Conference on Articial Intelligence (IJCAI), pp. 1–7.

Kongsro, J., 2014. Estimation of pig weight using a Microsoft Kinect prototype imaging system. Comput. Electron. Agric. 109, 32–35. https://doi.org/10.1016/j.compag.2014.08.008.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521 (7553), 436–444. https://doi.org/10.1038/nature14539.

Lin, T.-Y., Goyal, P., Girshick, R., He, K., Dollár, P., 2017. In: Focal Loss for Dense Object Detection, https://doi.org/10.1016/j.ajodo.2005.02.022.

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., et al., 2014. Microsoft COCO: common objects in context. In: ecture Notes in Computer Science, pp. 740–755.

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., Berg, A.C., 2016. SSD: single shot MultiBox detector. In: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). LNCS, pp. 21–37.

Madsen, T.N., Kristensen, A.R., 2005. A model for monitoring the condition of young pigs by their drinking behaviour. Comput. Electron. Agric. 48 (2), 138–154. https://doi.org/10.1016/j.compag.2005.02.014.

Mallick, T., Das, P.P., Majumdar, A.K., 2014. Characterizations of noise in kinetic depth images: a review. IEEE Sens. J. 14 (6), 1731–1740. https://doi.org/10.1109/JSEN.2014.2309987.

Manning, C.D., Schütze, H., 1999. Foundations of Statistical Natural Language Processing. MIT Press.

Maselyne, J., Saeys, W., Briene, P., Mertens, K., Vangeyte, J., De Ketelaere, B., et al., 2016. Methods to construct feeding visits from RFID registrations of growing-finishing pigs at the feed trough. Comput. Electron. Agric. 128, 9–19. https://doi.org/10.1016/j.compag.2016.08.010.

Maselyne, J., Van Nuffel, A., Briene, P., Vangeyte, J., De Ketelaere, B., Millet, S., et al., 2017. Online warning systems for individual fattening pigs based on their feeding pattern. Biosyst. Eng. 1–14. https://doi.org/10.1016/j.biosystemseng.2017.08.006.

Matthews, S.G., Miller, A.L., Clapp, J., Plötz, T., Kyriazakis, I., 2016. Early detection of health and welfare compromises through automated detection of behavioural changes in pigs. Vet. J. 217, 43–51. https://doi.org/10.1016/j.tvjl.2016.09.005.

Matthews, S.G., Miller, A.L., Plötz, T., Kyriazakis, I., 2017. Automated tracking to measure behavioural changes in pigs for health and welfare monitoring. Sci. Rep. 7 (1), 1–12. https://doi.org/10.1038/s41598-017-17451-6.

McFarlane, N.J.B., Schofield, C.P., 1995. Segmentation and tracking of piglets in images. Mach. Vis. Appl. 8 (3), 187–193. https://doi.org/10.1007/BF01215814.

Nasirahmadi, A., Edwards, S.A., Matheson, S.M., Sturm, B., 2017a. Using automated image analysis in pig behavioural research: assessment of the influence of enrichment substrate provision on lying behaviour. Appl. Anim. Behav. Sci. 196 (February), 30–35. https://doi.org/10.1016/j.applanim.2017.06.015.

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

Nasirahmadi, A., Hensel, O., Edwards, S.A., Sturm, B., 2016. Automatic detection of mounting behaviours among pigs using image analysis. Comput. Electron. Agric. 124, 295–302. https://doi.org/10.1016/j.compag.2016.04.022.

Nasirahmadi, A., Richter, U., Hensel, O., Edwards, S., Sturm, B., 2015. Using machine vision for investigation of changes in pig group lying patterns. Comput. Electron. Agric. 119, 184–190. https://doi.org/10.1016/j.compag.2015.10.023.

Navarro Jover, J.M.M., Alcañiz-Raya, M., Gómez, V., Balasch, S., Moreno, J.R.R., Grau Colomer, V., Torres, A., 2009. An automatic colour-based computer vision algorithm for tracking the position of piglets. Spanish J. Agric. Res. 7 (3), 535. https://doi.org/10.5424/sjar/2009073-438.

Nilsson, M., Herlin, A.H., Ardö, H., Guzhva, O., Aström, K., Bergsten, C., et al., 2015. Development of automatic surveillance of animal behaviour and welfare using image analysis and machine learned segmentation technique. Animal 9 (11), 1859–1865. https://doi.org/10.1017/S1751731115001342.

Pezzuolo, A., Guarino, M., Sartori, L., González, L.A., Marinello, F., 2018. On-barn pig weight estimation based on body measurements by a Kinect v1 depth camera. Comput. Electron. Agric. 148 (March), 29–36. https://doi.org/10.1016/j.compag.2018.03.003.

Poggio, T., Kawaguchi, K., Liao, Q., Miranda, B., Rosasco, L., Boix, X., et al., 2017. Theory of Deep Learning III: Explaining the Non-overfitting Puzzle. Retrieved from. http://arxiv.org/abs/1801.00173.

Redmon, J., Farhadi, A., 2018. YOLOv3: an incremental improvement. ArXiv 1–6.

Ren, S., He, K., Girshick, R., Sun, J., 2015. Faster R-CNN: towards real-time object detection with region proposal networks. IEEE Trans. Pattern Anal. Mach. Intell. 39 (6), 1137–1149. https://doi.org/10.1109/TPAMI.2016.2577031.

Schmidhuber, J., 2015. Deep learning in neural networks: an overview. Neural Networks 61, 85–117. https://doi.org/10.1016/j.neunet.2014.09.003.

Srivastava, N., Hinton, G.E., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. (JMLR) 15, 1929–1958.

Vranken, E., Berckmans, D., 2017. Precision livestock farming for pigs. Anim. Front. 7 (1), 32. https://doi.org/10.2527/af.2017.0106.

Weary, D.M., Huzzey, J.M., Von Keyserlingk, M.A.G., 2009. Board-invited review: using behavior to predict and identify ill health in animals. J. Anim. Sci. 87 (2), 770–777. https://doi.org/10.2527/jas.2008-1297.

Wolpert, D.H., 1996. The existence of a priori distinctions between learning algorithms. Neural Comput. 8 (7), 1391–1420. https://doi.org/10.1162/neco.1996.8.7.1391.

Yang, Q., Xiao, D., Lin, S., 2018. Feeding behavior recognition for group-housed pigs with the faster R-CNN. Comput. Electron. Agric. 155 (October), 453–460. https://doi.org/10.1016/j.compag.2018.11.002.

Zheng, C., Zhu, X., Yang, X., Wang, L., Tu, S., Xue, Y., 2018. Automatic recognition of lactating sow postures from depth images by deep learning detector. Comput. Electron. Agric. 147, 51–63. https://doi.org/10.1016/j.compag.2018.01.023.

Zoph, B., Shlens, J., Vasudevan, V., Shlens, J., Le, Q.V., Vasudevan, V., et al., 2017. Learning transferable architectures for scalable image recognition. ArXiv Preprint https://doi.org/1707.07012.