Journal of Agriculture and Food Research 16 (2024) 101067

Contents lists available at ScienceDirect

Journal of Agriculture and Food Research

journal homepage: www.sciencedirect.com/journal/journal-of-agriculture-and-food-research

Pose estimation of sow and piglets during free farrowing using deep learning

Fahimeh Farahnakian [a,*], Farshad Farahnakian [a], Stefan Björkman [b], Victor Bloch [c], Matti Pastell [c], Jukka Heikkonen [a]

[a] Department of Computing, University of Turku, Turku, 20500, Finland
[b] Department of Production Animal Medicine, University of Helsinki, 00014, Finland
[c] Resources Institute Finland (Luke), Latokartanonkaari 9, Helsinki, 00790, Finland

ARTICLE INFO

Keywords:
Deep learning
Convolutional neural networks
Livestock
Pose estimation
Animal behavior

ABSTRACT

Automatic and real-time pose estimation is important in monitoring animal behavior, health, and welfare. In this paper, we utilized pose estimation for monitoring the farrowing process to prevent piglet mortality and preserve the health and welfare of the sow. State-of-the-art Deep Learning (DL) methods have lately been used for animal pose estimation. This paper aims to probe the generalization ability of five common DL networks (ResNet50, ResNet101, MobileNet, EfficientNet, and DLCRNet) for sow and piglet pose estimation. These architectures predict the body parts of several piglets and the sow directly from input video sequences. Real farrowing data from a commercial farm was used for training and validation of the proposed networks. The experimental results demonstrated that MobileNet was able to detect seven body parts of the sow with a median test error of 0.61 pixels.

1. Introduction

Worldwide, pork production is expected to increase tremendously within the next decades [1,2] which will challenge the economic sustainability of pig producers and the welfare and health of the animals [3]. The challenge is to meet both the growing animal health demands as well as the growing pork demands without exploitation of the environment [4]. Further, farmers need to enhance animal health to not only increase revenue but also to decrease the use of medication such as antibiotics, and secure meat quality for the consumer. Unfortunately, the more intense the production the more challenging it is to preserve animal health and welfare. For instance, highly productive sows have longer farrowing durations, higher risk for puerperal disease, e.g., PostPartum Dysgalactia Syndrome (PPDS), and shorter longevity, and their litters are at higher risk of disease and mortality [5–7]. To reduce these diseases of sows and the mortality of piglets, farmers would need sufficient time for monitoring the parturition of these animals. In real life, because the profit margin per individual animal is low, the available time for farm workers to attend the parturition of individual sows and their litters is insufficient, which makes it more difficult to monitor and manage parturition correctly [1]. This is concerning for both the health and welfare of the animals as well as the economics of the pig producers. Piglets dying before weaning can result in economic losses between €12 and €23 per litter [5]. The economic losses caused by PPDS can reach between €300 and €470 per affected sow [5]. Therefore, intensive pig production with a high number of piglet-producing sows, high litter size, and long parturition are of economic, welfare, and environmental concern. Farmers need to manage an increasingly demanding situation in balancing production costs and loss of revenues against financial returns. It is important to provide tools and support for pig farmers to produce pork sustainably and profitably, and to meet welfare demands at the same time [3].

Assessing the behavior of livestock is important because it is a cheap and non-invasive way to get information about animal health and welfare [1]. However, manual monitoring, documentation, and assessment of animal behavior are difficult and inefficient because of the large number of animals and time-dependent physiological changes in behavior around parturition [8,9]. Therefore, reliable technologies to assist farmers in continuous monitoring and interpretation of the health of individual animals are needed in future pig production [10]. Precision livestock farming (PLF) is used to optimize farming processes and reduce human workload [11]. PLF provides possibilities for the farmers to

* Corresponding author.
E-mail address: farfar@utu.fi (F. Farahnakian).

https://doi.org/10.1016/j.jafr.2024.101067
Received 20 December 2023; Received in revised form 11 February 2024; Accepted 21 February 2024
Available online 19 March 2024

monitor individual animals despite the increase in the number of animals per farm [12]. According to Refs. [1,2,11], PLF will help farmers in continuous monitoring of their animals which will improve their health and welfare at any given time. As mentioned above, parturition is a critical event in pig production when it comes to improving the survival and health of the animals as well as the economics and sustainability of the producers. Several different PLF tools are now available that can be used at parturition to replace the farmer's eyes and ears in monitoring individual animals [1,2,10]. Therefore, PLF is an excellent alternative to human observation in the monitoring of animal behavior around parturition.

Automatic intelligent monitoring systems provide an efficient way to continuously analyze animal behavior to detect anomalies and enhance animal care. Animal behavior can be monitored with wearable sensors such as radio-frequency identification ear tags [13], Global Positioning System (GPS) [14] and Inertial Measurement Unit (IMU) [15]. However, the major drawback of using these sensors is installation and maintenance for each animal which bothers both human operators and animals. Therefore, sensors attached to the animal can disrupt the normal behavior of the animal and can malfunction because they can get broken or contaminated due to the laying of moving of the animal inside the farrowing pen. Video-based monitoring systems are used as a practical and convenient solution to address this problem [16]. As the systems use single or multiple cameras to capture images, they are also cost-effective due to cheaper hardware. Therefore, analysis of videos based on computer vision is currently considered the best alternative because of a robust and non-contact sensor that allows for continuous monitoring and analyzing of animal behavior [17]. In these works [8,18], they used depth images for automatic posture change analysis of lactating sows.

Pose Estimation (PE) plays a key role in intelligent systems for animal behavior analysis. It predicts the position of body parts of an animal and extracts information for different inspection purposes such as animal feeding [19], drinking [20], animal interaction [21], and tracking of movement [22]. PE for multiple animals can be more difficult as it needs to understand the number of animals, recognize and group the key points of each animal, and explore the position of each key point belonging to one animal. The multi-object PE can be classified into bottom-up methods and top-down methods [23]. In bottom-up methods (Fig. 1 (a)), first all key points of each animal's body parts are detected. After that, the poses of the animals are generated by grouping all parts belonging to distinct animals. In Ref. [24], the possibility of using DeepLabCut has been investigated for the pig PE task and a bottom-up method has been used. In top-down methods (Fig. 1 (a)), animals are first detected by using animal detectors to obtain the bounding box of each animal. Then, single-animal pose estimators are employed for each bounding box to produce multi-animal poses. In Ref. [25], an advanced multi-person PE framework using a top-down method has been proposed. The calculating time in top-down methods depends on the number of objects. Furthermore, the speed of estimating animal poses in the bottom-up method is faster than in the top-down method as there is no need to separately detect poses for each animal. Hence, the Bottom-up method is used in this paper to predict key points and estimate pig poses.

In recent years, Deep Learning (DL) techniques have made huge progress for animal PE [26], object detection [27], image classification [28], and face recognition [29]. For instance, Convolutional Neural Network (CNN) as a main DL model has been used in pig PE [8,30,31]. Although the proposed CNN-based models could show good

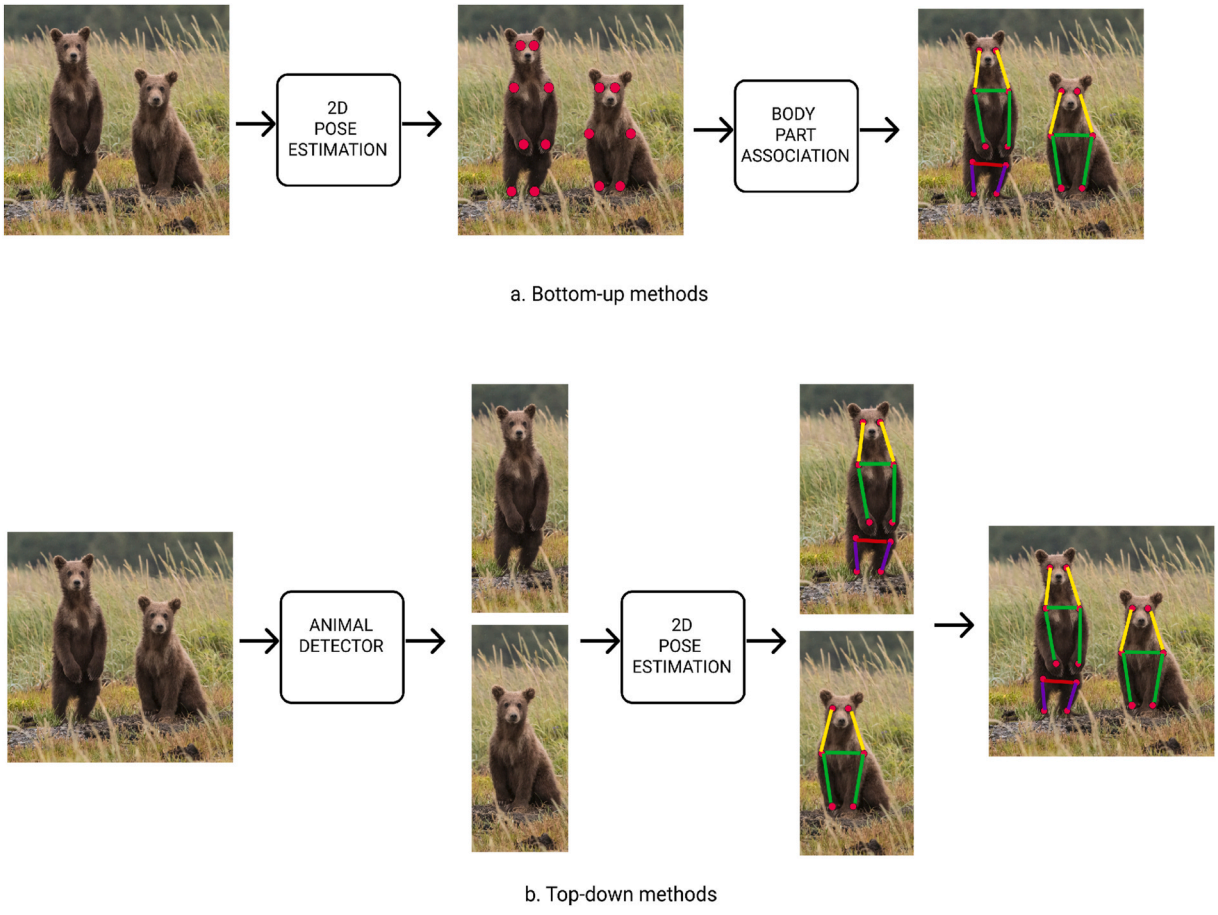


Fig. 1. Multi-animal 2D PE frameworks. (a) In bottom-up methods, all key points need to be detected (1). These key points are then associated with each animal and assembled for individual pose representations (2). (b) Top-down methods have two sub-tasks: (1) animal detection and (2) PE in the region of a single animal.

performance in finding correlations and eventually the condition of animals, it is still a challenging task because of the lack of a clean background, light changes, and object deformation. Further, for predicting multiple animal positions, overcrowded scenes are problematic because occlusion occurs when several animals are too close and combine with each other. In Ref. [24], we proposed a ResNet50 [32] model for the PE of individual pigs in environments where several pigs are present in one pen. We got the results on open source data and we annotated 2000 images of pigs from different locations and light conditions. As we found the ability of the open-source toolbox in Ref. [24], DeepLabCut [33], for the pig PE, we used the same toolbox here for developing PE methods on our collected data.

In this paper, we propose a CNN-based PE method to predict piglets and body parts of piglets in a sequence of images (video) without markers. A total of seven key points were manually annotated in each image for the location of the left leg, right leg, left hoof, right hoof, shoulder, tail, and snout of the sow. For the piglets, we only annotated the shoulder. We investigated the performance of five popular deep networks including ResNet50, ResNet101, MobileNet, EfficientNet and DLCRNet for PE of sow and piglet(s). To our knowledge, this is the first study using PE for both sow and piglets during the farrowing process using DeepLabCut.

The remainder of the paper is organized as follows. The proposed DL-based PE methods are described in Section 2 followed by the experimental setup in Section ?? and experimental results in Section 3. Finally, the conclusion is presented in 5.

2. Methods and materials

In this section, we review five well-known CNN-based algorithms that have been studied in this paper for PE. Furthermore, we provide information about the dataset and present implementation details for the evaluation of the proposed PE networks.

2.1. An overview of proposed deep networks

Deep learning for PE: PE methods are categorized into two separate groups: 2D and 3D [23]. The 2D spatial position of desired points is computed from the digital videos or images. 3D singular-object PE methods have the potential to obtain 3D pose annotations of an animal or human body. Estimating a 2D singular-object position is much easier because advanced lab environments need to have a 3D version of body key points. The 2D PE methods are classified into single-object and multi-object [34]. The single-object PE is applied when there is only a single object in the input image. Single-object PE methods can be classified into regression-based and heatmap-based. Fig. 2 shows the framework of these two types of methods. Regression-based methods (Fig. 2 (a)) apply DL-based pose regressor to learn a mapping from an input image to body key points for generating joint coordinates [35]. Heatmap-based methods (Fig. 2 (b)) focused on estimating the probability of the existing key points in each pixel of an input image [36]. Therefore, the key body points will not be directly detected from an input image.

ResNet or Residual Network is a supervised, feed-forward deep neural network model [32]. It uses layers as learning activation functions which reference to the input layers, instead of using learning functions without any reference. In other words, it allows to establishment and training of neural networks consisting of a large number (even thousands) layers with a low percentage of training error. Due to the structure of ResNet, it is able to tackle the vanishing gradient problem [37] caused by the activation function during the training network with gradient-based learning methods. A block of a ResNet network including two convolutional layers is shown in Fig. 3, where x is entered as an input to a first convolution layer with the residual function $f(x)$, and x simultaneously is added to the output of second convolution layer ($f(x) + x$) to pass to next blocks.

MobileNet is a particular family of CNNs with the capability of applying to mobile, internet-connected devices, and embedded systems

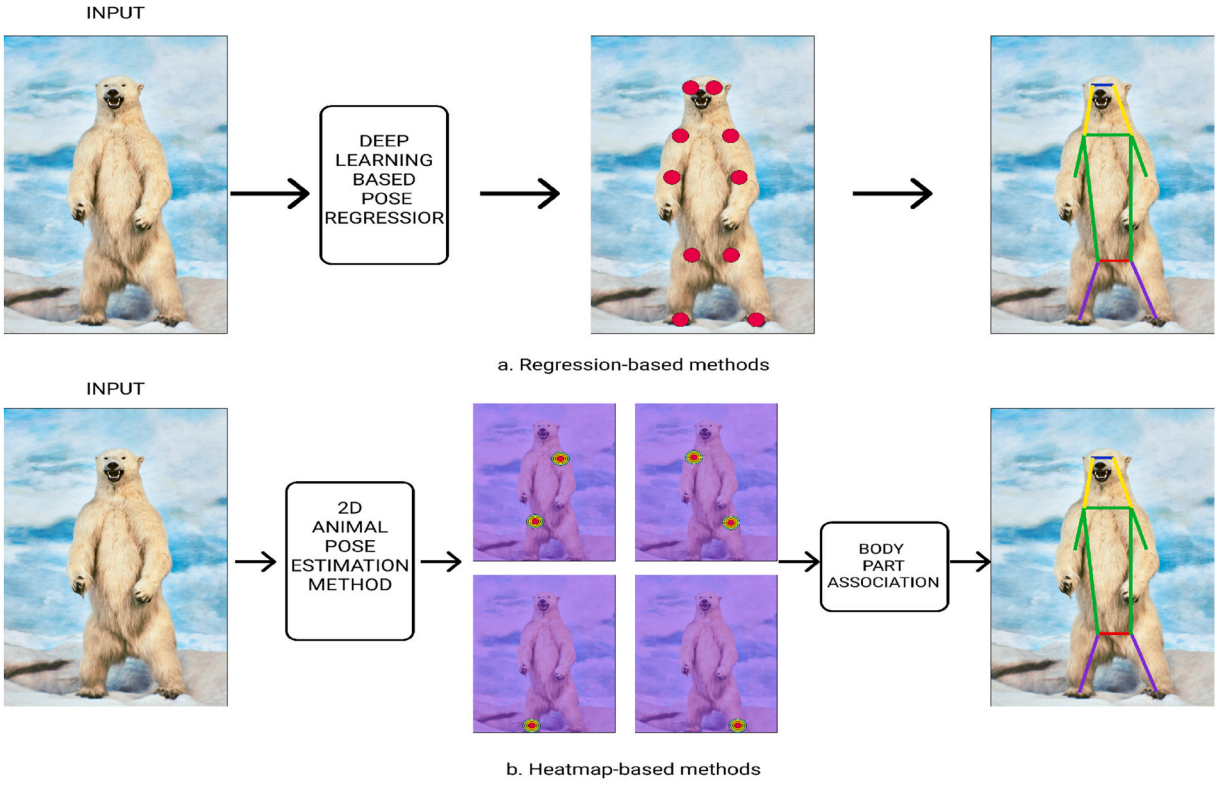


Fig. 2. Single-animal 2D PE frameworks. (a) Regression-based methods generate joint coordinates by directly learning a mapping from the source image to the kinematic body model (through a deep neural network). (b) Heatmap-based methods detect the position of body parts of animals using the supervision of heatmaps.

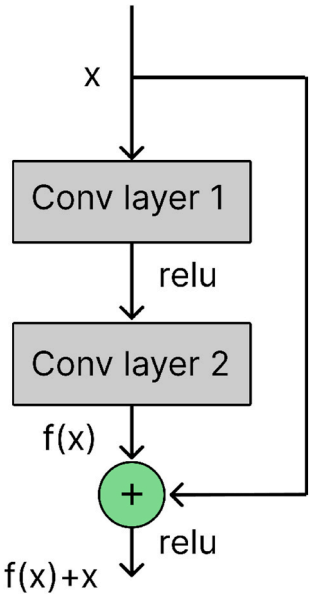


Fig. 3. A block of a ResNet network.

to do computer vision tasks like object detection, face detection, and logo or text recognition [38]. Furthermore, it has increasingly been used in the field of PE, for instance, model [39] which was created based on the MobileNet architecture has been proposed to estimate human poses. MobileNet model is constructed to be utilized in mobile applications because of its small model size (low number of parameters) and because they are less complex compared to other models as fewer multiplications and additions are used in this kind of model. Therefore, the mentioned features help to have more accuracy, less needed memory, and minor time consumption [40]. Regarding to MobileNet structure, it is observed that a standard convolution layer has not been used in its skeleton, and it uses depthwise separable convolutions [40], which notably decreases the number of parameters. The computing process in depth-wise separable convolutions is performed by depth-wise and point-wise convolutions. First, a filter is applied to each input channel by depth-wise convolution. A point-wise convolution layer then combines all output from the depth-wise convolution by applying a 1*1 convolution. In other words, the depth-wise convolution uses two separated layers for both filtering and combining processes, and this is exactly the difference between MobileNet and a standard CNN (see Fig. 4).

EfficientNet: Scaling up a convolutional network is a common way to get better accuracy on bench-marking datasets and to improve model quality. For instance, ResNet200 has achieved better accuracy than ResNet18 by just adding more layers [41]. The problem is that most of the techniques (width-wise, depth-wise, and image resolution) used for scaling up convolutional networks are picked randomly. As a consequence, the process of scaling up requires manual tuning and it is time-consuming. EfficientNet [42] is an impressive and easy scaling technique that helps to scale up any CNN model more systematically to address the problem mentioned above. As shown in Fig. 5, to construct EfficientNet, each scaling technique from width-wise to resolution scaling has been studied in Ref. [42] and they have noticed that balancing three dimensions (width, depth, and image resolution) with a fixed set of coefficient affected model's performance positively.

DLCRNet: There are some cases of DL-based architectures used in computer vision tasks, in which they can not show great performances. For instance, when these methods face digital images including objects of variable size and scale, or when they have to perform visual tasks at the pixel level. In these circumstances, deep learning-based techniques such as Faster-RCNN and YOLO are unsuccessful since visual features from small areas vanish during convolutional and pooling processes.

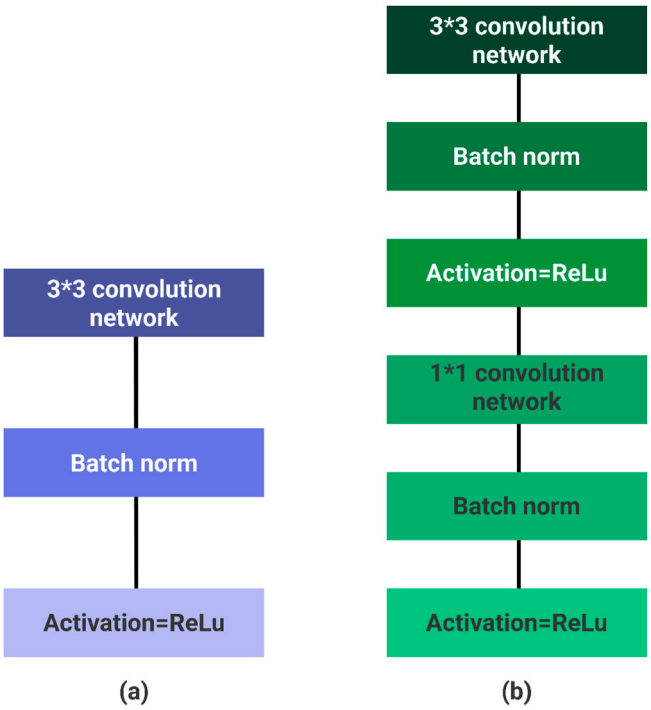


Fig. 4. (a) Standard convolutional network with Relu and batch norm. (b) Depth-wise separable convolution with depth-wise and pointwise layers which are followed by batch normalization and Relu.

One of the solutions that are recently used in animal PE to extract features in any resolutions and scales is to employ a CNN-based multi-scale network architecture where the information from high-resolution feature maps (small scale) integrates with information from low-resolution feature maps (large scale) by neural networks [43,44]. In general, multi-scale network architectures are classified into three groups: multi-column network [45], skip-net [46], and multi-scale input [47]. In the multi-column network shown in Fig. 6(a), input data are fed into various columns. The output data of each parallel column are then interconnected as the final output. In Fig. 6(b), the skip-net connects low-scale features with a large-scale output. Thus, features of distinct scales are composed and fed into an output layer. In the multi-scale input method illustrated in Fig. 6(c), the input images are divided into several scales. DLCRNet [33] is a multi-scale input architecture, and the input of this model is the fusion of both low- and high-resolution feature maps. This ability of the DLCRNet method reduces missing key points and helps the PE system to recognize key points of animal bodies in each scale level [44].

2.2. Dataset and implementation details

Data was collected on a private family piggery located at Oripää (Finland) during 4–6.2021. The videos were recorded by three 5Mpixel IP cameras (HDBW5541R-ASE-0280B, Dahua, China) with 2.8 mm lenses (102° × 71° FOV), which were attached to water supplying pipes at a height of about 2 m above the floor of free farrowing pens (Fig. 7). The cameras were connected to a PC (16 Gb, AMD Ryzen 7) with video recording software (BlueIris) and farrowing was recorded and files stored on an external 4 TB memory disk (Seagate).

A video of 16 farrowing was recorded. The videos were saved starting 2 h before the farrowing till 5 h after the end of the farrowing. This period is the most critical for preventing piglet mortality and preserving the health of the sow. The recording duration per farrowing was 12–32 h, totaling 310 h. The video was recorded in different lighting conditions: from daylight with direct sunlight passing through windows till twilight with only red light from the infrared light in the piglet nest.

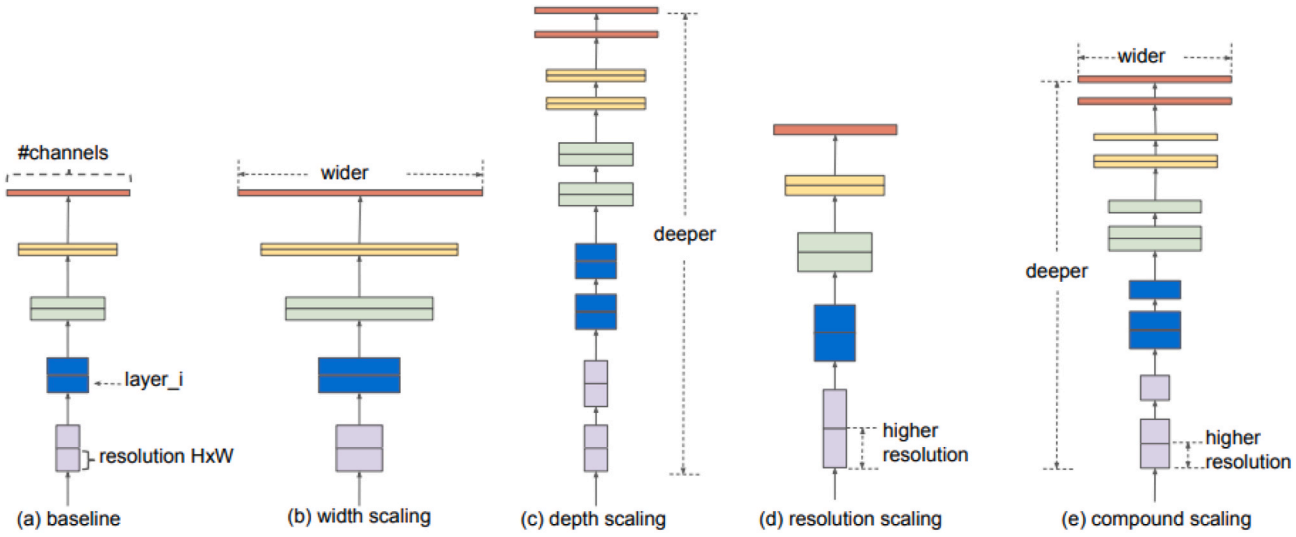


Fig. 5. Various scaling methods and compound scaling [42].

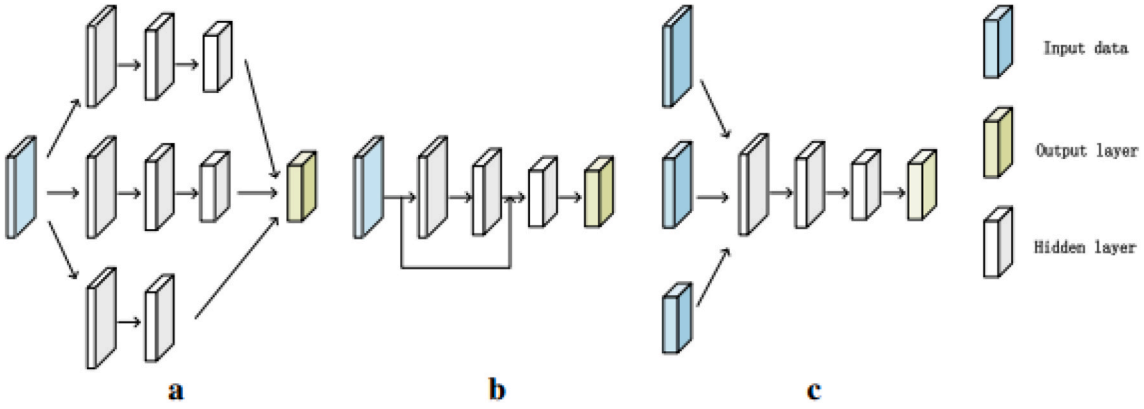


Fig. 6. a) multi-column network, b) skip-net, c) multi-scale input [48].

Fig. 7. Setup of the recording camera.(a) Side view and (b) Camera view.

In this study, we selected 856 frames generated from the recorded videos to represent different time points of the farrowing process and different light conditions. Each image consists of a sow and a different number of piglets. The maximum number of piglets in an image was 18. Hence, a total maximum of 19 objects were in an image including sow.

In this study, we undertook a precise manual annotation process, focusing on key anatomical landmarks to facilitate accurate PE for both sows and piglets. For the sows, which are considerably larger and feature more discernible body parts in the video footage, we annotated seven critical points: the shoulder, left leg, right leg, left hoof, right hoof, tail, and snout. These points were selected as they represent pivotal joints and extremities that are essential for constructing a comprehensive skeletal model of the sow's posture, enabling us to capture a wide range of movements and behaviors relevant to the farrowing process. Given the smaller size of piglets, associated with the limitations imposed by the resolution of our video data, we were constrained to annotate only a single point—the shoulder—for each piglet. This decision was driven by the practical challenge of accurately distinguishing and marking multiple distinct points on piglets within the crowded and dynamically changing environment of the farrowing pen. The shoulder point serves as a proxy for the piglet's position relative to the Sow, which, despite its simplicity, provides valuable insights into the spatial dynamics critical for assessing aspects of welfare and risk during early life stages. Fig. 8 presents an exemplar image from our dataset, illustrating the annotated points on both a sow and her piglets.

DeepLabCut toolbox (version 2.2.2) [33] has been used for body part PE. DeepLabCut consists of a markerless technique that was invented to extract detailed human and animal poses without using any markers in locations with dynamic backgrounds. We evaluated five networks including ResNet-50, ResNet-101, MobileNet, EfficientNet, and DLCRNet which are included in the DeepLabCut toolbox. We trained all networks with 60000 training iterations with batch size 8. It has to be mentioned that DeepLabCut supports automatic hyperparameter tuning to monitor model performance and make informed decisions about hyperparameter adjustments. In this study, we used the automatic hyperparameter tuning feature. For instance, the batch size or the number of frames used per training iteration suggested by the automatic feature was 8.

We used a random split to divide the dataset. 80% of the dataset served for training and the other 20% served for testing. Further, we used image augmentation to generate more images in order to create a robust and accurate model. In Ref. [49], they provided the experimental results to show augmentation is a promising solution for improving DL models' performance if the augmentation methods are chosen in a proper way and they do not affect the semantic information in the image. For this purpose, we augmented the training dataset with a random transformation including blur, image rotation, flip, crops, and contrast change. Post-augmentation, our dataset expanded significantly, providing a diverse range of images for training and effectively mitigating the risk of overfitting. The final augmented dataset comprised 1027 frames, ensuring comprehensive coverage of different scenarios. An example image with different augmentation methods is illustrated in Fig. 9.

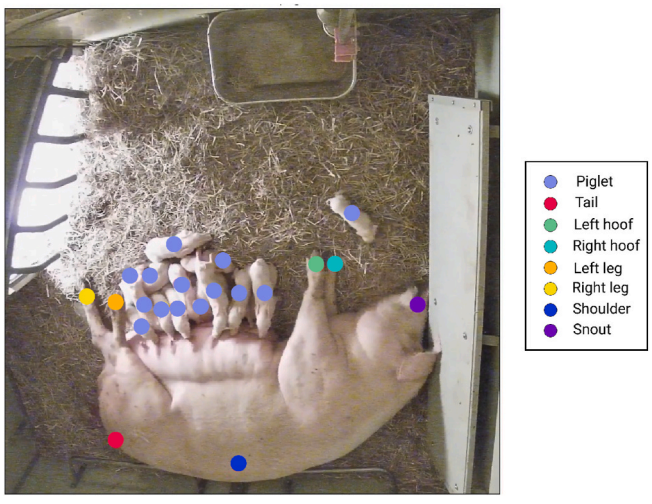


Fig. 8. Example of an annotated image from our dataset with different colors corresponding to the body part of the sow and piglets.

To prevent potential overfitting due to the high number of training iterations, we employed two common strategies: Early Stopping [50] technique and Adam regularization method [51] in all networks. During the training process, the early stopping could stop training when the validation loss does not improve for a predefined number of iterations to protect against overfitting.

3. Results

We computed the error and the root mean square error for evaluation of the performance of the five proposed networks. In addition, the score and location refinement maps are also computed for further analysis of the results. Finally, the results of body part detection are illustrated for five networks on three images from the test dataset.

3.1. Root mean squared error (RMSE)

We calculated the RMSE between the location of the detected point and the reference point in pixels for each frame and key point in the test dataset. RMSE curves for the training and testing process for each network have been illustrated in Fig. 10. The training RMSE begins at a higher value, showing a natural decline as the network models learn and adapt to the training data. Concurrently, the testing RMSE—initially higher than the training RMSE—follows a similar downward trajectory, yet it remains above the training RMSE for most of the training duration. This persistent gap underscores the intrinsic challenge of achieving strong generalization to unseen data. Particularly, the plots reveal that MobileNet not only reaches lower error rates more expediently but also maintains a tighter convergence between the training and testing errors, suggesting a superior ability to generalize compared to the other evaluated networks.

Additionally, the RMSE curves for ResNet-50 and ResNet-101 incorporate early stopping at epochs 250 and 300, respectively. These points are denoted by green dashed lines, signifying the moments when the training process is halted to prevent overfitting. Post these epochs, the RMSE values plateau, indicating that no further learning occurs and reinforcing the effectiveness of early stopping in safeguarding the models' ability to generalize.

Fig. 11(a) shows the RMSE of sow's body parts for the five proposed networks. The results show that all networks performed well for body part detection. The minimum RMSE belongs to the shoulder part of the sow. In DLCRNet, RMSE was 0.39, 0.95, 0.93, 0.89, 0.89, 0.48, and 0.77 pixels for shoulder, left leg, right leg, left hoof, right hoof, tail, and snout, respectively. The median test errors of ResNet-50, ResNet-101, MobileNet, EfficientNet, and DLCRNet were 0.93, 0.97, 0.61, 0.87, and 0.89, respectively. Therefore, the performance of MobileNet is better than other models. Moreover, We obtained the RMSE for the one annotated body part of the piglets (shoulder). Fig. 11(b) illustrates this result for the five networks in the test dataset. It shows that the minimum error was achieved by MobileNet. However, values of RMSE for the piglet's shoulder in all networks are very similar.

3.2. Error

Error was used to evaluate the distance in pixels between the manual labels and the predicted ones for both the training and test datasets. We also calculated this value with a defined parameter that is called "p-

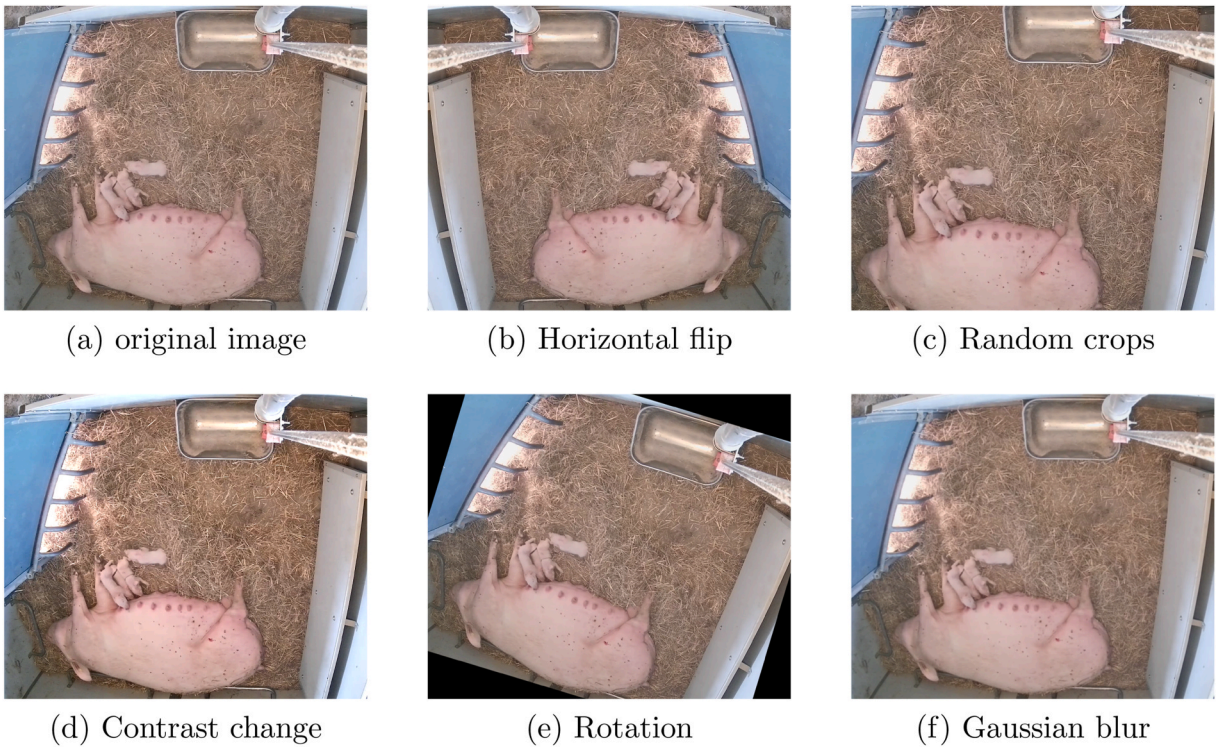


Fig. 9. The proposed augmentation techniques on an example image from the dataset.

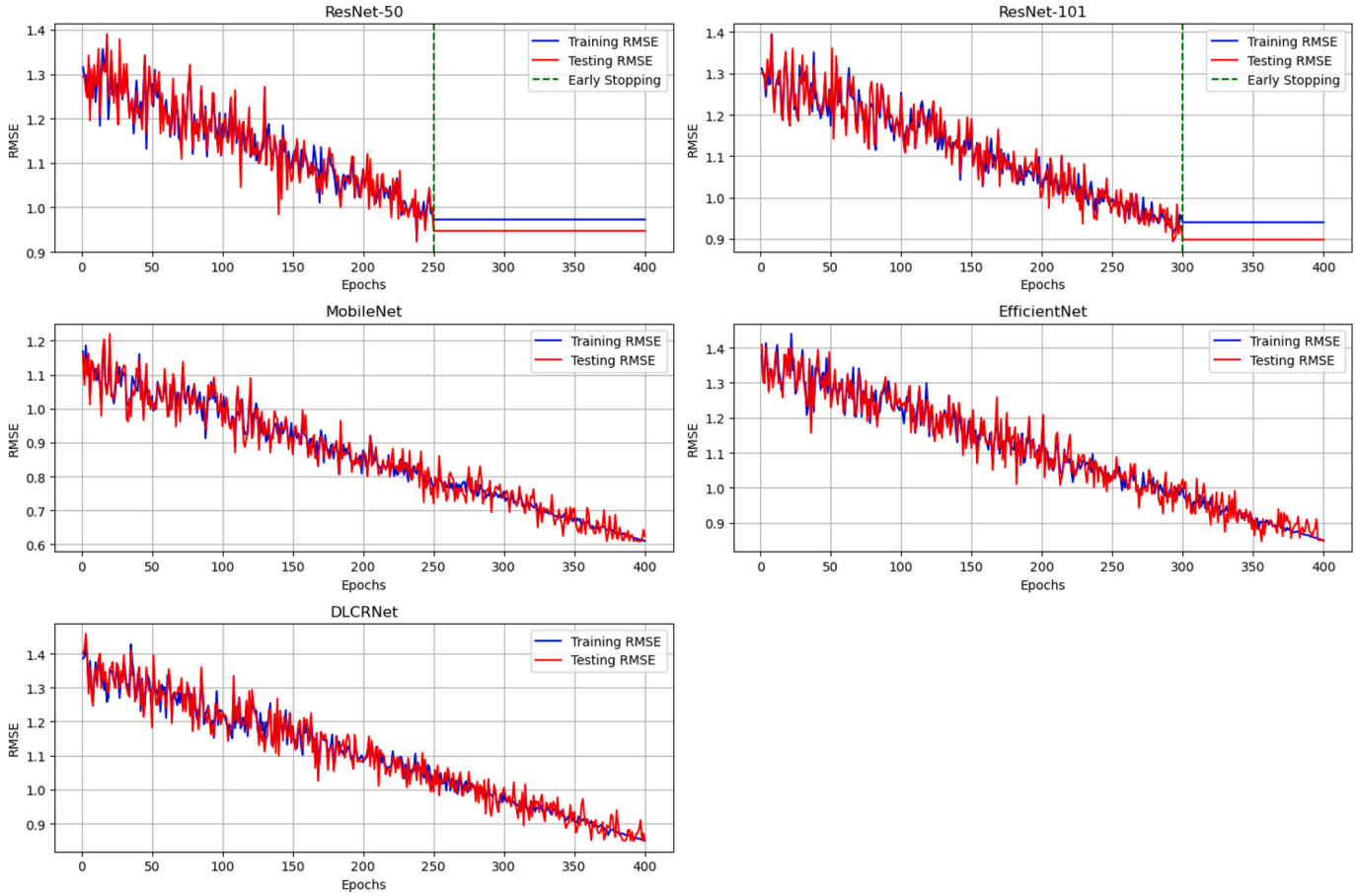


Fig. 10. RMSE curves for the training and testing processes for each network (ResNet-50, ResNet-101, MobileNet, EfficientNet, and DLCRNet).

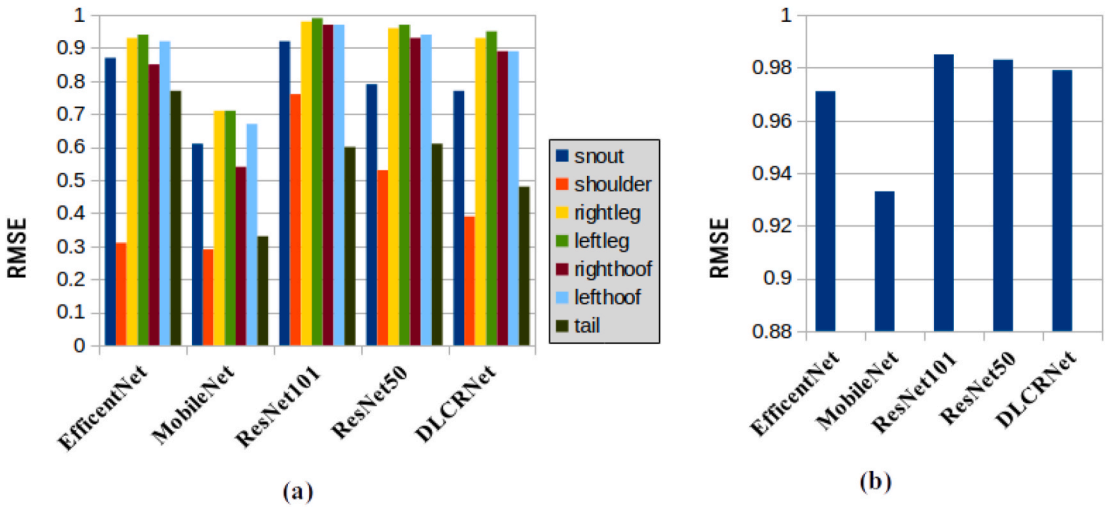


Fig. 11. The RMSE of (a) sow's body parts and (b) piglet's shoulder in pixels for each network.

cutoff". The value of this parameter is set to 0.6, meaning that predictions with a likelihood of <0.6 will be displayed as uncertain predictions. So "train error with p-cutoff" means how often the training data did not reach the p-cutoff. Table 1 shows these error values for all networks. The results show that ResNet101 got the minimum train error compared to other networks. However, the minimum test error was achieved by MobileNet. The minimum test and train error with p-cutoff were obtained from MobileNet and ResNet101, respectively.

3.3. Qualitative results

Fig. 12 demonstrates three test dataset examples for each proposed network. Dots represent the manual annotations whereas '+' represents the high-confidence predictions. In Fig. 12, the distance between the labeled (dots) and the predicted body parts ('+') is small. The same color code in Fig. 8 is assumed for body parts in this image. The results show that these models can perform PE with high accuracy (good generalization). However, there are a few false detections.

3.4. Score and location refinement maps

Score and location refinement maps were calculated based on ground truth data of annotated key points. Score maps show the probability that a key point occurs at a particular location. Location refinement maps predict offsets to mitigate quantization errors due to downsampled score maps [33,52]. From the result in Figs. 13 and 14, these maps can easily attend to any category of image and correctly identify the selected area. For instance, the red areas have higher attention scores as shown in Fig. 14. In summary, all small piglets are detected in three images, especially the one that is close to the pen's wall (top right of images).

Table 1
Train and test error (with p-cutoff) of the proposed networks in pixels.

	EfficientNet	MobileNet	ResNet101	ResNet50	DLCRNet
Train error	12.71	16.28	5.92	6.97	8.12
Test error	34.98	24.12	28.71	37.21	33.48
Train error with p-cutoff	9.12	12.2	4.72	4.96	5.04
Test error with p-cutoff	19.95	16.72	18.6	19.12	21.66

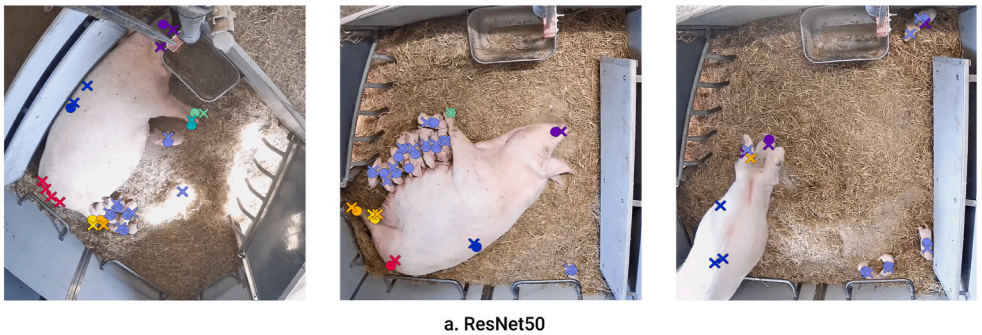
4. Discussion

Our study contributes to the growing body of research on PE in animal welfare and behavior monitoring, leveraging state-of-the-art deep learning architectures within the DeepLabCut framework. Unlike traditional methods that may rely on object detection algorithms like YOLOv3 [53] and Faster R-CNN [54] for identifying animal positions, our approach focuses on precise PE through the identification and grouping of key body points. This distinction is crucial in the context of monitoring complex behaviors and interactions within the farrowing pen, where the precise positioning and movement of sows and piglets are of paramount importance.

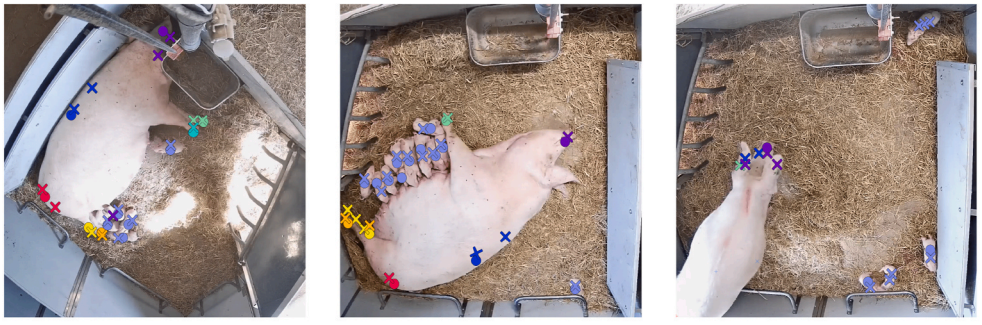
Furthermore, the significance of monitoring maternal behavior to enhance piglet survival and sow welfare cannot be overstated. Our study's emphasis on detailed PE of sows and, to a lesser extent, piglets, seeks to build upon the foundation laid by prior research in the use of 2D [55–57] and 3D imaging systems [8,18,58] for behavior analysis. While previous studies have successfully utilized various sensor types and camera technologies for monitoring a range of behaviors [59–61], or monitoring posture and postural change of sows [8,18,56,58] and tracking of individual piglets [55] in a farrowing and lactation pen, our work extends these efforts by applying advanced deep-learning techniques to offer more nuanced insights into PE.

The choice of technology and analytical tools plays a critical role in the effectiveness of PE methods. Although MATLAB has been widely used in earlier studies for image processing and algorithm development [18,56,58], our application of DeepLabCut introduces the advantages of a dedicated PE tool, capable of handling the complexities associated with animal behavior analysis. This choice reflects our commitment to adopting and adapting the latest technological advancements to address specific research needs.

Our experimental evaluation revealed significant insights into the performance of various deep-learning networks for sow and piglet PE. The analysis, grounded in both quantitative metrics such as RMSE and qualitative assessment through error rates, underscores the capabilities of each network in the context of precision livestock farming. The RMSE results provide a clear indication of the precision with which each network could identify the body parts of sows and the shoulder point of piglets. Notably, MobileNet emerged as the standout performer, demonstrating the lowest median test error among the evaluated networks. For further investigation, we also calculated the percentage improvement in the performance of each network relative to the baseline (ResNet-101, which has the highest RMSE) by using the following formula:

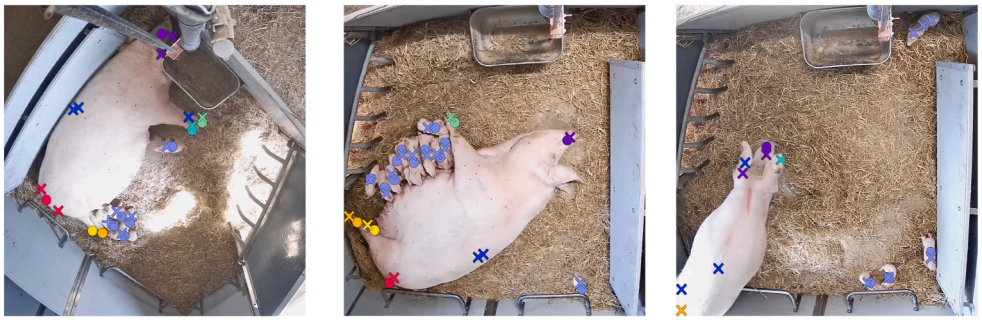


a. ResNet50



b. ResNet101

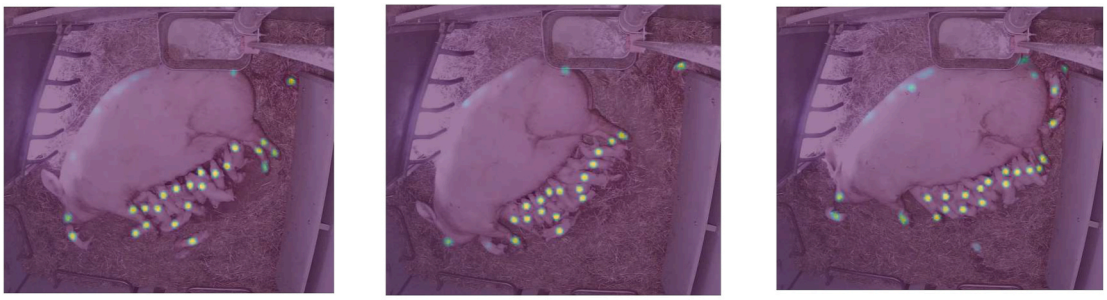
c. MobileNet



d. EfficientNet

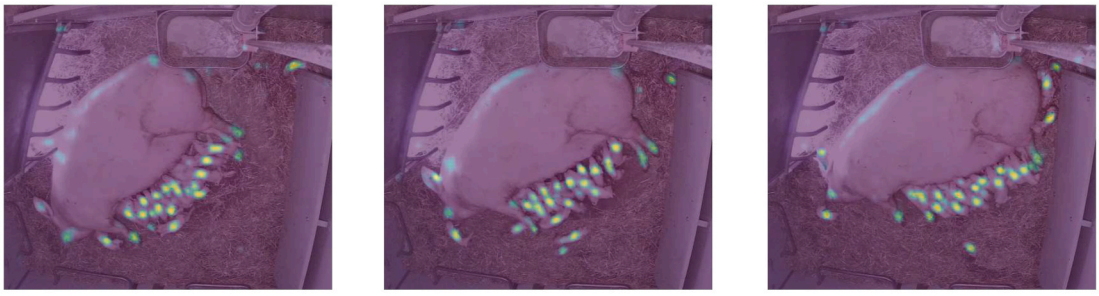
e. DLCRNet

Fig. 12. Qualitative results of the proposed networks on three examples of the test dataset. The human labels are plotted as a dot and the framework's predictions are plotted as a plus symbol ('+').

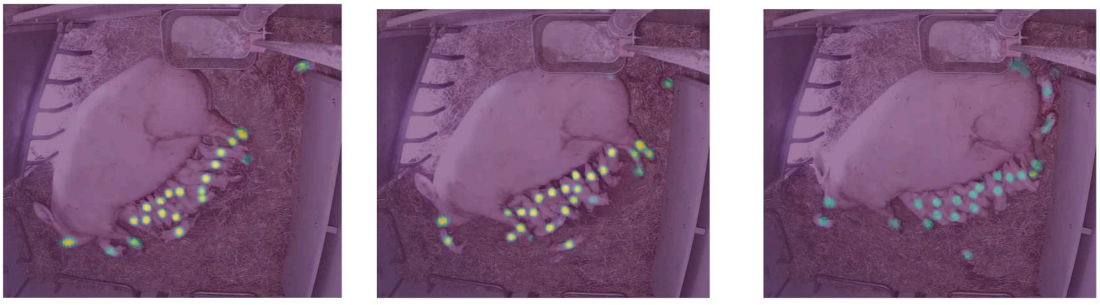


a. ResNet50

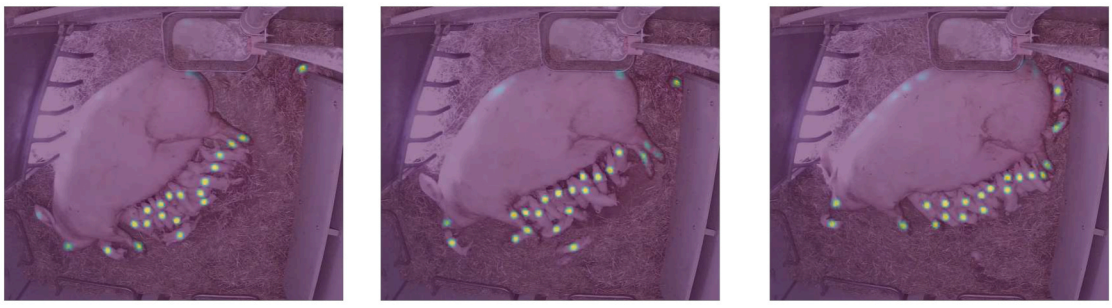
b. ResNet101



c. MobileNet



d. EfficientNet



e. DLCRNet

Fig. 13. The comparison of the score maps of the proposed networks.

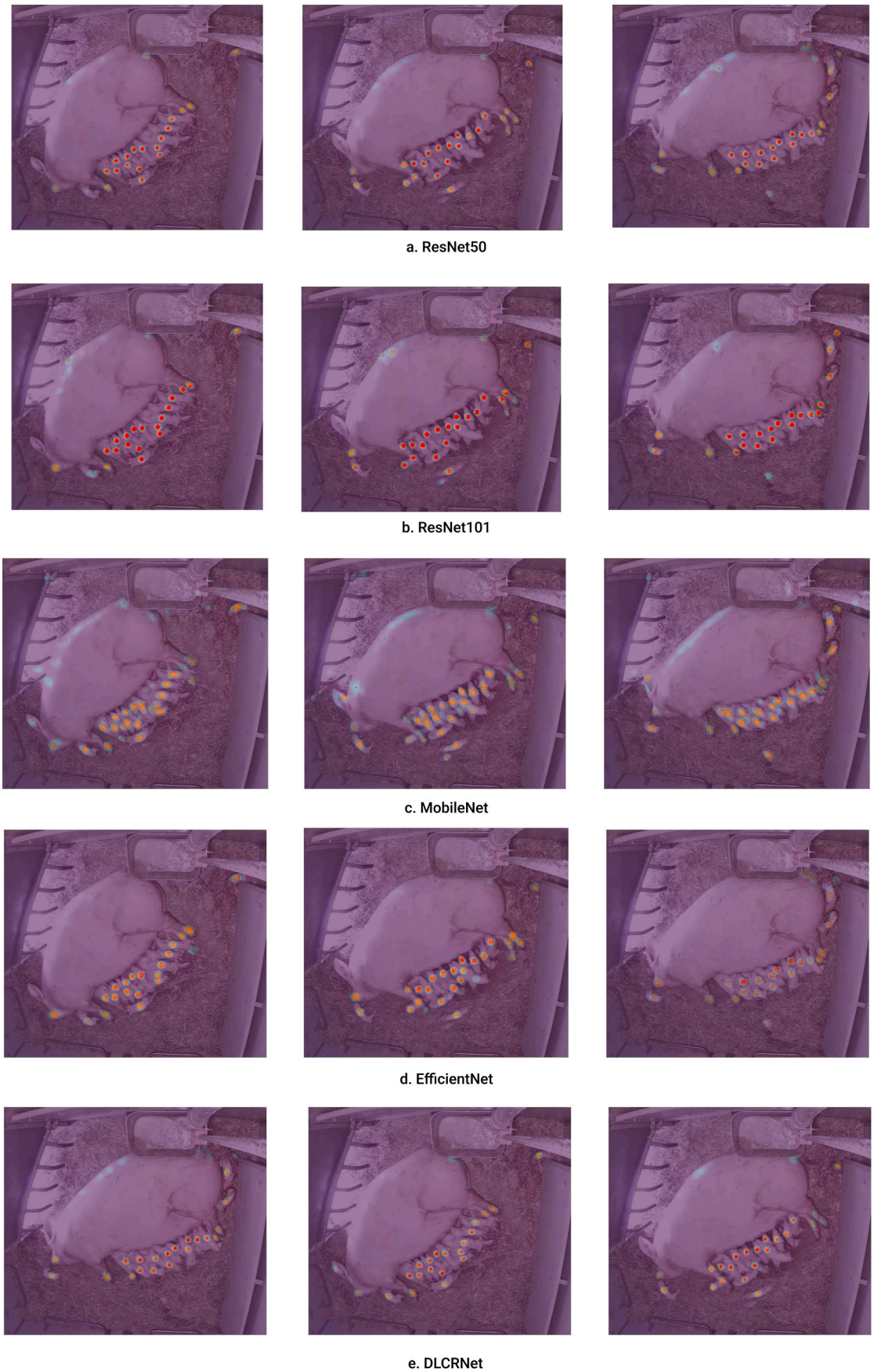


Fig. 14. The comparison of the location refinement maps of the proposed networks.

$$\text{Percentage Improvement} = \left(\frac{\text{RMSE\_baseline} - \text{RMSE\_model}}{\text{RMSE\_baseline}}\right) \times 100\%$$

According to the obtained results, ResNet-50 demonstrated a modest improvement of 4.12%, while EfficientNet and DLCRNet showed improvements of 10.31% and 8.25%, respectively. Particularly, MobileNet achieved a substantial performance leap, registering a 37.11% improvement over the baseline model, ResNet-101, which itself showed no improvement as it served as the reference point for our comparison. These findings underscore MobileNet's ability to accurately predict poses with high efficiency compared to other networks again.

However, our study also acknowledges the limitations inherent in the current scope of PE. The limitations encountered in our study, particularly concerning the detailed PE of piglets, highlight the need for continuous innovation in camera and imaging technologies. The potential integration of additional cameras focused specifically on piglets, as well as the exploration of higher-resolution imaging solutions, represents a promising avenue for future research. Such advancements will enable more comprehensive monitoring of all individuals within the farrowing pen, further contributing to our understanding and enhancement of animal welfare. Furthermore, we recognize that an expanded set of criteria could encompass a broader spectrum of behaviors and interactions, potentially uncovering more insights into the dynamics between sows and piglets. To tackle this, our future plans include augmenting the number of frames, taking into account additional factors such as: (1) Various times of the day to account for changes in activity levels and lighting conditions. (2) A range of piglet behaviors beyond the immediate farrowing period, including feeding habits, social interactions, and reactions to environmental stimuli. (3) The integration of a wider variety of environmental contexts within the farrowing pen, like the closeness to heating elements and the amount of available space.

5. Conclusion

We predicted different body parts of the sow and piglets during farrowing. We investigated the performance of five deep learning-based networks for PE via the powerful DeepLabCut toolbox. The obtained results on our collected data set show that the proposed methods can predict the body parts of both sow and piglet(s) efficiently. The developed method can be used to monitor and estimate the pose of the sow and her piglets during the parturition. Therefore, it can be used to reduce unnecessary disturbance of the animals which will most likely improve the welfare and health of the animals as well as the economics and profitability of pig farmers. In addition, this work will lead from routine to evidence-based parturition management which will have an additional positive effect on piglet survival and health of the sow.

For future work, we will gather more input videos by using different sensors. For instance, by fusing lidar and infrared camera data, we may be able to get a more robust system that can monitor during different light conditions, especially at nighttime. Further, based on the PE information, we plan to classify the behavior of the sow into sitting, lying, standing, and walking. We will also define sudden changes in the posture of the sow, e.g., rolling. Tracking the piglets in space and time will allow for recognition, e.g., isolation of single piglets and synchronization of piglet behavior within the litter.

Funding

This study was funded by the Finnish Ministry of Agriculture (grant decision 529/03.01.02/2018) and further financially supported by the Algorithmic Computational Intelligence research group at the University of Turku.

CRediT authorship contribution statement

Fahimeh Farahnakian: Software, Methodology. **Farshad Farahnakian:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology. **Stefan Björkman:** Writing – review & editing, Writing – original draft, Resources, Project administration, Data curation. **Victor Bloch:** Writing – review & editing, Writing – original draft, Resources, Data curation. **Matti Pastell:** Writing – review & editing, Writing – original draft, Resources, Data curation. **Jukka Heikkonen:** Validation, Supervision.

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.

Data availability

The data that has been used is confidential.

Acknowledgments

The authors wish to acknowledge CSC – IT Center for Science, Finland, for computational resources. Further, we would like to thank the owners of the private family farm in Oripää/Finland: Petri Matintalo and Karin Dahl.

References

[1] D. Berckmans, Precision livestock farming technologies for welfare management in intensive livestock systems, Rev Sci Tech 33 (1) (2014) 189–196, https://doi.org/10.20506/rst.33.1.2273.
[2] D. Berckmans, General introduction to precision livestock farming, Animal Frontiers 7 (1) (2017) 6–11, https://doi.org/10.2527/af.2017.0102.
[3] A. Malak-Rawlikowska, M. Gębska, R. Hoste, C. Leeb, C. Montanari, M. Wallace, K. de Roest, Developing a methodology for aggregated assessment of the economic sustainability of pig farms, Energies 14 (6) (2021), https://doi.org/10.3390/en14061760.
[4] C. D.S, M. S.N, B. Kemp, Recent advances in pig reproduction: focus on impact of genetic selection for female fertility, Reproduction in domestic animals = Zuchthygiene 33 (1) (2018) 28–36, https://doi.org/10.1111/rda.13264.
[5] M.-L. iemi JK, P.H.M. Bergman, Modeling the costs of postpartum dysgalactia syndrome and locomotory disorders on sow productivity and replacement, Front. Vet. Sci. 4 (181) (2017), https://doi.org/10.3389/fvets.2017.00181.
[6] E. Baxter, S. Edwards, Piglet Mortality and Morbidity: Inevitable or Unacceptable? Elsevier, Netherlands, 2018, pp. 73–100, https://doi.org/10.1016/B978-0-08-101012-9.00003-4.
[7] C. Oliviero, S. Junnikkala, O. Peltoniemi, The challenge of large litters on the immune system of the sow and the piglets, Reprod. Domest. Anim. 54 (53) (2019) 12–21, https://doi.org/10.1111/rda.13463.
[8] C. Zheng, X. Zhu, X. Yang, L. Wang, S. Tu, Y. Xue, Automatic recognition of lactating sow postures from depth images by deep learning detector, Comput. Electron. Agric. 147 (2018) 51–63, https://doi.org/10.1016/j.compag.2018.01.023. https://www.sciencedirect.com/science/article/pii/S0168169917309985.
[9] O. Peltoniemi, C. Oliviero, J. Yun, A. Grahofer, S. Björkman, Management practices to optimize the parturition process in the hyperprolific sow, J. Anim. Sci. 98 (2020) S96–S106, https://doi.org/10.1093/jas/skaa140.
[10] E. Vranken, D. Berckmans, Precision livestock farming for pigs, Animal Frontiers 7 (1) (2017) 32–37, https://doi.org/10.2527/af.2017.0106.
[11] Y. Gómez, A. Stygar, I. Boumans, E. Bokkers, L. Pedersen, J. Niemi, M. Pastell, X. Manteca, P. Llonch, A systematic review on validated precision livestock farming technologies for pig production and its potential to assess animal welfare, Frontiers in Veterinary Science 8, publisher Copyright: © Copyright © 2021 Gómez, Stygar, Boumans, Bokkers, Pedersen, Niemi, Pastell, Manteca and Llonch (May 2021), https://doi.org/10.3389/fvets.2021.660565.
[12] C. Tzanidakis, P. Simitzis, K. Arvanitis, P. Panagakis, An overview of the current trends in precision pig farming technologies, Livest. Sci. 249 (2021) 104530, https://doi.org/10.1016/j.livsci.2021.104530. https://www.sciencedirect.com/science/article/pii/S1871141321001384.
[13] J. Maselyne, I. Adriaens, T. Huybrechts, B. De Ketelaere, S. Millet, J. Vangeyte, A. Van Nuffel, W. Saeys, Measuring the drinking behaviour of individual pigs housed in group using radio frequency identification (rfid), Animal : an international journal of animal bioscience 11 (2015) 1–10, https://doi.org/10.1017/S1751731115000774.
[14] I. Pray, D. Swanson, V. Ayvar, C. Muro, L.m. Moyano, A. Gonzalez, H.H. Garcia, S. O'Neal, Gps tracking of free-ranging pigs to evaluate ring strategies for the control of cysticercosis/taeniasis in Peru, PLoS Neglected Trop. Dis. 10 (2016) e0004591, https://doi.org/10.1371/journal.pntd.0004591.

[15] H.J. Escalante, S.V. Rodriguez, J. Cordero, A.R. Kristensen, C. Cornou, Sow-activity classification from acceleration patterns: a machine learning approach, Comput. Electron. Agric. 93 (2013) 17–26, https://doi.org/10.1016/j.compag.2013.01.003. http://www.sciencedirect.com/science/article/pii/S0168169913000082.
[16] C. Zheng, X. Yang, X. Zhu, C. Chen, L. Wang, S. Tu, A. Yang, Y. Xue, Automatic posture change analysis of lactating sows by action localisation and tube optimisation from untrimmed depth videos, Biosyst. Eng. 194 (2020) 227–250, https://doi.org/10.1016/j.biosystemseng.2020.04.005. https://www.sciencedirect.com/science/article/pii/S1537511020300945.
[17] J. Bao, Q. Xie, Artificial intelligence in animal farming: a systematic literature review, J. Clean. Prod. 331 (2022) 129956, https://doi.org/10.1016/j.jclepro.2021.129956. https://www.sciencedirect.com/science/article/pii/S0959652621041251.
[18] F. Lao, T. Brown-Brandl, J. Stinn, K. Liu, G. Teng, H. Xin, Automatic recognition of lactating sow behaviors through depth image processing, Comput. Electron. Agric. 125 (2016) 56–62, https://doi.org/10.1016/j.compag.2016.04.026. https://www.sciencedirect.com/science/article/pii/S0168169916301612.
[19] T. Brown-Brandl, G. Rohrer, R. Eigenberg, Analysis of feeding behavior of group housed growing–finishing pigs, Comput. Electron. Agric. 96 (2013) 246–252, https://doi.org/10.1016/j.compag.2013.06.002. http://www.sciencedirect.com/science/article/pii/S0168169913001324.
[20] M. Kashiha, C. Bahr, S.A. Haredasht, S. Ott, C.P. Moons, T.A. Niewold, F.O. Ödberg, D. Berckmans, The automatic monitoring of pigs water use by cameras, Comput. Electron. Agric. 90 (2013) 164–169, https://doi.org/10.1016/j.compag.2012.09.015. http://www.sciencedirect.com/science/article/pii/S0168169912002372.
[21] S. Viazzi, G. Ismayilova, M. Oczak, L. Sonoda, M. Fels, M. Guarino, E. Vranken, J. Hartung, C. Bahr, D. Berckmans, Image feature extraction for classification of aggressive interactions among pigs, Comput. Electron. Agric. 104 (2014) 57–62, https://doi.org/10.1016/j.compag.2014.03.010. http://www.sciencedirect.com/science/article/pii/S0168169914000805.
[22] M.A. Kashiha, C. Bahr, S. Ott, C.P. Moons, T.A. Niewold, F. Tuyttens, D. Berckmans, Automatic monitoring of pig locomotion using image analysis, Livest. Sci. 159 (2014) 141–148, https://doi.org/10.1016/j.livsci.2013.11.007. http://www.sciencedirect.com/science/article/pii/S1871141313005003.
[23] C. Zheng, W. Wu, C. Chen, T. Yang, S. Zhu, J. Shen, N. Kehtarnavaz, M. Shah, Deep Learning-Based Human Pose Estimation: A Survey, 2020, https://doi.org/10.48550/ARXIV.2012.13392.
[24] F. Farahnakian, J. Heikkonen, S. Björkman, Multi-pig pose estimation using deeplabcut, in: 2021 11th International Conference on Intelligent Control and Information Processing (ICICIP), 2021, pp. 143–148, https://doi.org/10.1109/ICICIP53388.2021.9642168.
[25] H.-S. Fang, S. Xie, Y.-W. Tai, C. Lu, Rmpe: regional multi-person pose estimation, in: 2017 IEEE International Conference on Computer Vision (ICCV), 2017, pp. 2353–2362, https://doi.org/10.1109/ICCV.2017.256.
[26] L. Liu, J. Zhou, B. Zhang, S. Dai, M. Shen, Visual detection on posture transformation characteristics of sows in late gestation based on libra r-cnn, Biosyst. Eng. 223 (2022) 219–231, https://doi.org/10.1016/j.biosystemseng.2022.09.003. https://www.sciencedirect.com/science/article/pii/S1537511022002070.
[27] F. Farahnakian, J. Heikkonen, Deep learning based multi-modal fusion architectures for maritime vessel detection, Rem. Sens. 12 (16) (2020), https://doi.org/10.3390/rs12162509.
[28] N. Jmour, S. Zayen, A. Abdelkrim, Convolutional neural networks for image classification, in: 2018 International Conference on Advanced Systems and Electric Technologies, 2018, pp. 397–402, https://doi.org/10.1109/ASET.2018.8379889.
[29] K. Yan, S. Huang, Y. Song, W. Liu, N. Fan, Face recognition based on convolution neural network, in: 2017 36th Chinese Control Conference (CCC), 2017, pp. 4077–4081, https://doi.org/10.23919/ChiCC.2017.8027997.
[30] J. Brünger, M. Gentz, I. Traulsen, R. Koch, Panoptic segmentation of individual pigs for posture recognition, Sensors 20 (13) (2020), https://doi.org/10.3390/s20133710.
[31] A. Nasirahmadi, B. Sturm, S. Edwards, K.-H. Jeppsson, A.-C. Olsson, S. Müller, O. Hensel, Deep learning and machine vision approaches for posture detection of individual pigs, Sensors 19 (17) (2019). URL, https://www.mdpi.com/1424-8220/19/17/3738.
[32] K. He, X. Zhang, S. Ren, J. Sun, Deep Residual Learning for Image Recognition, 2015, https://doi.org/10.48550/ARXIV.1512.03385.
[33] A. Mathis, P. Mamidanna, K. Cury, T. Abe, V. Murthy, M. Mathis, M. Bethge, Deeplabcut: markerless pose estimation of user-defined body parts with deep learning, Nat. Neurosci. 21 (9) (2018), https://doi.org/10.1038/s41593-018-0209-y.
[34] P. Jafarzadeh, P. Virjonen, P. Nevalainen, F. Farahnakian, J. Heikkonen, Pose estimation of hurdles athletes using openpose, in: 2021 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME), 2021, pp. 1–6, https://doi.org/10.1109/ICECCME52200.2021.9591066.
[35] A. Toshev, C. Szegedy, DeepPose: human pose estimation via deep neural networks, in: 2014 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2014, https://doi.org/10.1109/cvpr.2014.214.
[36] Q. Dang, J. Yin, B. Wang, W. Zheng, Deep learning based 2d human pose estimation: a survey, Tsinghua Sci. Technol. 24 (6) (2019) 663–676, https://doi.org/10.26599/TST.2018.9010100.
[37] S. Basodi, C. Ji, H. Zhang, Y. Pan, Gradient amplification: an efficient way to train deep neural networks, Big Data Mining and Analytics 3 (3) (2020) 196–207, https://doi.org/10.26599/BDMA.2020.9020004.
[38] A.G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, H. Adam, Mobilenets: Efficient Convolutional Neural Networks for Mobile Vision Applications, 2017, https://doi.org/10.48550/ARXIV.1704.04861.
[39] B. Debnath, M. O'Brien, M. Yamaguchi, A. Behera, Adapting mobilenets for mobile based upper body pose estimation, in: 2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), 2018, pp. 1–6, https://doi.org/10.1109/AVSS.2018.8639378.
[40] F. Chollet, Xception: Deep Learning with Depthwise Separable Convolutions, 2016, https://doi.org/10.48550/ARXIV.1610.02357.
[41] Y. Huang, Y. Cheng, A. Bapna, O. Firat, M.X. Chen, D. Chen, H. Lee, J. Ngiam, Q. V. Le, Y. Wu, Z. Chen, Gpipe: Efficient Training of Giant Neural Networks Using Pipeline Parallelism, 2018, https://doi.org/10.48550/ARXIV.1811.06965.
[42] M. Tan, Q.V. Le, Efficientnet: Rethinking Model Scaling for Convolutional Neural Networks, 2019, https://doi.org/10.48550/ARXIV.1905.11946.
[43] L.-C. Chen, Y. Yang, J. Wang, W. Xu, A.L. Yuille, Attention to scale: scale-aware semantic image segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.
[44] J. Lauer, M. Zhou, S. Ye, W. Menegas, T. Nath, M.M. Rahman, V. Di Santo, D. Soberanes, G. Feng, V.N. Murthy, G. Lauder, C. Dulac, M.W. Mathis, A. Mathis, Multi-animal pose estimation and tracking with deeplabcut, bioRxiv (2021), https://doi.org/10.1101/2021.04.30.442096 arXiv:https://www.biorxiv.org/content/early/2021/04/30/2021.04.30.442096.full.pdf.
[45] N. Neverova, C. Wolf, G.W. Taylor, F. Nebout, Multi-scale deep learning for gesture detection and localization, in: L. Agapito, M.M. Bronstein, C. Rother (Eds.), Computer Vision - ECCV 2014 Workshops, Springer International Publishing, Cham, 2015, pp. 474–490.
[46] D. Eigen, R. Fergus, Predicting depth, surface normals and semantic labels with a common multi-scale convolutional architecture, in: Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 2650–2658.
[47] C. Farabet, C. Couprie, L. Najman, Y. LeCun, Learning hierarchical features for scene labeling, IEEE Trans. Pattern Anal. Mach. Intell. 35 (8) (2013) 1915–1929, https://doi.org/10.1109/TPAMI.2012.231.
[48] L. Wang, B. Yin, A. Guo, H. Ma, J. Cao, Skip-connection convolutional neural network for still image crowd counting, Appl. Intell. 48 (10) (2018) 3360–3371, https://doi.org/10.1007/s10489-018-1150-1.
[49] A. Mathis, S. Schneider, J. Lauer, M.W. Mathis, A Primer on Motion Capture with Deep Learning: Principles, Pitfalls and Perspectives, CoRR abs/2009.00564, 2020. arXiv:2009.00564.
[50] L. Prechelt, Early stopping-but when?, in: Neural Networks: Tricks of the Trade Springer, 2002, pp. 55–69.
[51] D. Kingma, J. Ba, Adam: a method for stochastic optimization, in: International Conference on Learning Representations, vol. 12, 2014.
[52] E. Insafutdinov, L. Pishchulin, B. Andres, M. Andriluka, B. Schiele, Deepercut: a deeper, stronger, and faster multi-person pose estimation model, CoRR abs/1605 (2016) 03170 arXiv:1605.03170.
[53] J. Redmon, A. Farhadi, Yolo9000: better, faster, stronger, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 6517–6525, https://doi.org/10.1109/CVPR.2017.690.
[54] S. Ren, K. He, R.B. Girshick, J. Sun, in: C. Cortes, N.D. Lawrence, D.D. Lee, M. Sugiyama, R. Garnett (Eds.), Faster R-Cnn: towards Real-Time Object Detection with Region Proposal Networks, NIPS, 2015, pp. 91–99. URL, http://dblp.uni-trier.de/db/conf/nips/nips2015.html/#RenHGS15.
[55] M. Oczak, K. Maschat, D. Berckmans, E. Vranken, J. Baumgartner, Automatic estimation of number of piglets in a pen during farrowing, using image analysis, Biosyst. Eng. 151 (2016) 81–89, https://doi.org/10.1016/j.biosystemseng.2016.08.018. https://www.sciencedirect.com/science/article/pii/S1537511016303464.
[56] A. Yang, H. Huang, X. Zhu, X. Yang, C. Pengfei, L. Shimei, Y. Xue, Automatic recognition of sow nursing behaviour using deep learning-based segmentation and spatial and temporal features, Biosyst. Eng. 175 (2018) 133–145, https://doi.org/10.1016/j.biosystemseng.2018.09.011.
[57] N. P, M. C, S. B, T. I, S. Küster, Automatic behavior and posture detection of sows in loose farrowing pens based on 2d-video images, Frontiers in Animal Science 64 (2021).
[58] S. Leonard, H. Xin, T. Brown-Brandl, B. Ramirez, Development and application of an image acquisition system for characterizing sow behaviors in farrowing stalls, Comput. Electron. Agric. 163 (2019) 104866, https://doi.org/10.1016/j.compag.2019.104866. https://www.sciencedirect.com/science/article/pii/S0168169919305666.
[59] Y. Chung, H. Kim, H. Lee, D. Park, T. Jeon, H.-H.C. and, A cost-effective pigsty monitoring system based on a video sensor, KSII Transactions on Internet and Information Systems 8 (4) (2014) 1481–1498, https://doi.org/10.3837/tiis.2014.04.018.
[60] Sanne Ott, Christel Moons, Mohammadamin A. Kashiha, Claudia Bahr, Frank Tuyttens, Daniel Berckmans, Theo A. Niewold, Automated video analysis of pig activity at pen level highly correlates to human observations of behavioural activities, Livest. Sci. 160 (2014) 132–137, https://doi.org/10.1016/j.livsci.2013.12.011.
[61] Mohammad Amin Kashisha, Claudia Bahr, Sanne Ott, Christel Moons, Theo A. Niewold, Frank Tuyttens, Daniel Berckmans, Automatic monitoring of pig locomotion using image analysis, Livest. Sci. 159 (2014) 141–148, https://doi.org/10.1016/j.livsci.2013.11.007.