A Review of Posture Detection Methods for Pigs Using Deep Learning

Zhe Chen, Jisheng Lu and Haiyan Wang

Abstract: Analysis of pig posture is significant for improving the welfare and yield of captive pigs under different conditions. Detection of pig postures, such as standing, lateral lying, sternal lying, and sitting, can facilitate a comprehensive assessment of the psychological and physiological conditions of pigs, prediction of their abnormal or detrimental behavior, and evaluation of the farming conditions to improve pig welfare and yield. With the introduction of smart farming into the farming industry, effective and applicable posture detection methods become indispensable for realizing the above purposes in an intelligent and automatic manner. From early manual modeling to traditional machine vision, and then to deep learning, multifarious detection methods have been proposed to meet the practical demand. Posture detection methods based on deep learning show great superiority in terms of performance (such as accuracy, speed, and robustness) and feasibility (such as simplicity and universality) compared with most traditional methods. It is promising to popularize deep learning technology in actual commercial production on a large scale to automate pig posture monitoring. This review comprehensively introduces the data acquisition methods and sub-tasks for pig posture detection and their technological evolutionary processes, and also summarizes the application of mainstream deep learning models in pig posture detection. Finally, the limitations of current methods and the future directions for research will be discussed.

Keywords: pig posture detection; deep learning; machine vision

1. Introduction

Numerous studies have shown that the posture of pigs can serve as an important indicator of their psychological and physiological state and help predict their natural behaviors, which are directly related to their healthy welfare, thereby affecting the production value of pigs [1–3]. For instance, a review has summarized various tail postures in relation to the physical and emotional states as well as injury behaviors of pigs [4]. Moreover, different pig postures are usually a result of the influence of various external factors [5–7], which can generally be controlled by breeders. Considering the close association shown in Figure 1, accurate detection and analysis of pig posture patterns can help producers adjust and optimize the breeding scheme so as to improve pig welfare and commercial value [8].

Besides the application in the estimation of pig welfare, pig posture detection can also contribute to other purposes. In pig counting, posture filtration is performed at the final stage to ultimately ascertain the number of pigs that are indeed feeding, which has been demonstrated to successfully reduce misjudgment in contrast to previous methods that neglected posture information [9]. In pig body size measurement, the image frames of non-upright or non-straight posture would be eliminated by posture classification to improve the measurement accuracy and stability [10–12].
solutions, and explore the development prospects of this field.

Figure 1. The correlation between external factors, psychological and physiological state of pigs, pig postures, and pig welfare and production. (The external factors include breeding conditions, environmental parameters, social interaction between pigs within the same enclosure, and invasive human activities [5–7]).

However, in previous research on animal behavior and posture, pig posture and location were typically recorded through on-the-spot observation or monitoring records [13]. Similarly, in real production environments, experts or breeders performed these procedures to adjust breeding strategies accordingly [14]. However, manual pig posture detection is considerably time-consuming therefore leading to low efficiency in the assessment of pig welfare. Moreover, even sophisticated experts may have the problems such as subjectivity, inconsistency, and even incorrect differentiation between highly similar postures, such as sternal recumbency and ventral recumbency [15]. Therefore, manual detection of pig posture is not feasible for commercial farm construction. With the rapid development of artificial intelligence (AI) in recent years, the concept of smart farming has gradually gained popularity in pig farming for automatic and continuous monitoring of livestock’s behavioral patterns, such as pig posture [16,17]. Naturally, highly efficient and effective methods for pig posture detection are crucial for such smart farming construction [18].

Fortunately, various animal-attached sensors and cameras have been introduced into the research scenarios and practical pig farms, enabling the capture and monitoring of animal biomarkers, such as pig posture [19]. With the assistance of these types of equipment, the progressive development of automated posture detection methods has been contributing to satisfactory posture detection schemes and stimulating smart farming construction [20]. With the evolution from complex sensors to common cameras, from early manual modeling to traditional machine vision, and then to deep learning, a multitude of detection methods have been proposed. Among them, detection methods based on deep learning have shown superior performance and promising potential to be popularized in commercial production on a large scale.

This review will comprehensively introduce the data acquisition process and sub-tasks for pig posture detection as well as their technological evolutionary processes. Additionally, it will also summarize various advanced deep learning methods applied in pig posture detection, with a primary focus on camera-assisted techniques. Furthermore, the review will examine the limitations of current methods, propose potential avenues for corresponding solutions, and explore the development prospects of this field.

2. Data Acquisition and Sub-Tasks for Pig Posture Detection
2.1. Data Acquisition for Pig Posture Detection
2.1.1. Non-Image Data from Animal-Attached Sensors

Before the maturity of image processing technologies, researchers generally employed various animal-attached sensors to obtain physical parameters related to pigs, and then
indirectly evaluated the behavior or posture patterns of pigs through data-specific modeling analysis [21]. Sensors were initially used to monitor the behavior and activity of individual captive pigs in research. Considering the close relationship between drinking behavior and the health welfare of pigs, Maselyne et al. [22] designed a high-frequency radio frequency identification (RFID) system that could automatically monitor and analyze the drinking behavior of individual pigs. Cornou et al. [23] used the acceleration data collected by a three-dimensional digital accelerometer to construct a multivariate dynamic linear model that can classify five types of activities and postures of sows in the farrowing house. By securing a tri-axial accelerometer to the hip of individual sows, three-dimensional motion data can be acquired and inputted to a Support Vector Machine (SVM) classifier for classification and quantification of the posture and posture transition of individuals [24]. Similarly, with a supervised machine learning approach, Escalante et al. [25] employed three-dimensional acceleration data from an accelerometer fitted on the neck of each pig to accomplish the classification of the five types of activities, including lateral and sternal lying.

Compared with manual observation, animal-attached sensors can automate most operations originally performed by laborers with the assistance of machine learning technologies and can produce more convincing and objective judgments [26]. However, there are still several disadvantages preventing their real application. First, pig posture detection with animal-attached sensors is cannot be universally used due to the variation of data type between different varieties of devices. Then, as intrusive data-gathering devices, sensors always need to be fixed in pig bodies, which may have some negative impacts on pig health [8]. Moreover, the fixation and accuracy of sensors are likely to be disturbed by other pigs and the sensors themselves [23], and the sensors may fail to work [24]. Therefore, there is a certain difficulty in the popularization of sensors under commercial farm conditions.

2.1.2. Two-Dimensional Images from Optical Cameras

With the rapid advancement of image processing techniques, ordinary 2D monitor cameras, which are characterized by convenience and inexpensiveness, have gradually become the most popular data acquisition equipment in animal behavior research [26]. Two-dimensional cameras used for pig posture detection, mainly including RGB cameras and gray cameras, are usually fixed above the pen with the lens pointing down to obtain a vertical top view [6]. Researchers can input certain images or a whole video sequence into a machine vision model to detect and estimate the individual posture of every pig within a pig pen in the pixel space.

Researchers have long been attempting to detect and analyze pig postures in a visual image for intuitive interpretation. With the help of image processing technologies, such as ellipse-fitting or segmentation algorithms, individual pigs in color images can be segmented and then detected under sub-optimal conditions [27], or changes in posture pattern and location of grouped pigs can be found in gray images [6]. Apart from the traditional methods, superabundant deep learning pig posture models based on the 2D camera system have been emerging in recent years. Application of deep learning can obtain considerably high average precision, above 98%, in pig behavior and posture identification at the individual level [28].

2.1.3. Depth Images from Depth Cameras

In addition to the 2D optical pixel information, the depth image can provide the auxiliary distance information between pigs and devices at every pixel available to detectors, thereby providing 3D data about the pigs. Therefore, the depth image can improve the poor performance of 2D camera systems under certain situations, such as occlusion among grouped pigs and difficulty in distinguishing sternal and ventral lying from the top view [21]. Along with providing additional depth information, depth images can reduce the negative influence of poor light conditions [29]. Depth images were initially applied to animal behavioral research, where the Kinect depth sensor showed excellent performance in recognizing aggressive behavior and automatic detection of lameness by analyzing the
walking patterns within depth images [30,31]. Like the 2D camera system, depth camera systems are also commonly used in pig posture detection. Lao et al. [32] used a top-view 3D Kinect camera to obtain depth images, which were then input into a manually designed recognition algorithm to recognize and classify pig behaviors and standing, sitting, and lying postures. In addition to the whole-body posture, to facilitate the prediction of tail biting in intensive pig farms, the posture or angle of pig tails can be easily measured by modeling analysis on depth images collected from time-of-flight 3D cameras [33].

Even though depth images are less sensitive to lighting conditions, their quality is easily affected by the occlusion of moving noises caused by the disturbance of dust and dirt [34]. Using the spatiotemporal interpolation technique to remove noise, Kim et al. [34] applied the background subtraction method to detect the standing and lying posture of pigs from depth images. In addition, depth images usually have lower resolution than traditional RGB images [29], which may decrease the quality of extracted features and then affect the detection of small or subtle changes in pig postures. This problem can be alleviated by powerful algorithm models. For example, Zheng et al. [29] used a deep learning model to process depth images acquired by Kinect v2 sensor, and Xu et al. [35] inputted depth image data obtained by Azure Kinect DK depth camera into a designed model that integrates deep learning techniques. In their experiments, five postures of pigs could be detected well and classified under commercial conditions.

2.2. Sub-Tasks for Image-Based Pig Posture Detection

Standard object detection refers to the discovery and location of target objects in a specific image and the concurrent identification of the category of every object [36]. As an extension of object detection, generic pig posture detection generally consists of pig localization and posture classification. Furthermore, in demanding application and research scenarios, identification of pig identity and timely tracking of pigs are included in this task for better practicability [28], and pig segmentation is included for better performance [37]. To more comprehensively introduce computer vision tasks for pig posture detection, this section decomposes pig posture detection into four technical submodules: pig localization, posture classification, pig segmentation, as well as pig identification and tracking, which will be explained in the following sections. Figure 2 shows the pipeline of utilizing the obtained raw data to apply to these four tasks. Particularly, the two fundamental issues of pig localization and posture classification, will be introduced in more detail for universality. Notably, pig localization and posture classification are executed in two separate stages in most traditional image processes and some deep learning methods, which are known as two-stage methods. However, they can be simultaneously completed in one stage in many state-of-the-art deep learning detectors.

2.2.1. Pig Localization

Pig localization refers to finding the precise position of all pigs and then either separating the contours of pigs from the background [35], framing them with bounding boxes [38], or generating a series of key points to represent their outline [39] in a given image or a specific video frame, which are exemplified in Figure 3 (The images used for this figure was sourced from the datasets of Psota [40], and the same applies to Figures 4–7). This is a fundamental issue in the computer vision field and also a precondition for pig posture detection.
Figure 2. The processes of utilizing acquired data to address real-world problems.

Figure 3. Examples of pig localization: (a) Example of pig localization by contours; (b) Example of pig localization by bounding boxes; (c) Example of pig localization by key points. (Blue marks indicate left and right neck, purple marks indicate left and right shoulder, green marks indicate left and right abdomen, red marks indicate left and right hip, and yellow marks indicate left and right tail [39]).

Figure 4. Example of pig posture classification.
different posture types are predicted simultaneously). Figure 6. A typical two-stage model pipeline for pig posture detection.

Figure 7. A typical one-stage model diagram for pig posture detection. (The image is divided into grids, and within each grid cell, bounding boxes, confidence scores, and class probabilities for different posture types are predicted simultaneously).
In early pig posture detection practices, pig localization was accomplished by extracting the binary contours of pigs from the background, which can be implemented by the threshold segmentation algorithm and watershed transformation [7,41]. However, posture detection with this scheme largely depends on the effect of segmentation algorithms since the contour quality has a direct influence on the feature extraction of contours, which will further influence the posture classification [42]. The sliding window technique once attracted great interest as a standard method in inchoate object detection models, such as pedestrian detection [43], which scans all possible locations in an image at different scales to determine the presence of any target within a window [44]. Using sliding windows and integrating principal component analysis convolution, Sun et al. [45] obtained a recall rate of up to 99.21% and a classification accuracy of 95.21% on pig detection when using 500 windows. The use of sliding windows can ensure that every area in an image can be noted by the detector, resulting in an omission detection rate and misdetection rate lower than those of You Only Look Once (YOLO) [46] and a detection speed faster than that of the Faster region convolutional neural network (Faster R-CNN) [47]. However, the sliding window also has the disadvantages of low computational efficiency and the reliance of window size and ratio on manual design.

To optimize the sliding window, a conception of region proposals [48] within the “recognition using regions” paradigm [49] was proposed in region-based convolutional neural networks (R-CNN) [48] series object detectors, which can selectively generate category-independent object proposals for the input image instead of scanning through the whole image. In R-CNN and Fast R-CNN, selective search [50] is used to generate object proposals that will be further estimated whether there is any object within the region by SVM. The most popular model in the R-CNN series used for pig posture detection is Faster R-CNN, which performs region proposals by a region proposal network (RPN) [47] and introduces the creation of anchor boxes that can generate high-quality region proposals cost-effectively. Application of Faster R-CNN to pig posture detection enables timely localization of pig position with a high average precision (AP) of 87.4% [51]. However, the R-CNN series models require some post-processing, such as bounding box regression used to refine the bounding boxes and re-correct the localization precision. Moreover, anchor boxes are also universally used in YOLO series models. In this case, these anchor boxes are evenly and densely distributed in a whole image, and their scale can be determined by the prior knowledge of ground truth in the training set. In recent years, methods based on anchor-free theory [52,53], such as key point detection for pigs [39], were gradually integrated into pig posture detection, which can localize the object in a novel manner and have no need for pre-set anchors, thereby improving the computational efficiency.

2.2.2. Posture Classification

Posture classification refers to the assignment of a label from several prepared posture labels (such as standing, sitting, and lying, and their descriptions are enumerated in Table 1) to individual pigs detected in an image, as shown in Figure 4. The procedures of posture classification mainly comprise feature-extraction engineering and classification based on the extracted feature vectors by a trained classifier [54].

<table>
<thead>
<tr>
<th>Posture Type</th>
<th>Description</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standing</td>
<td>Upright body position on extended legs with hooves only in contact with the floor</td>
<td>[29]</td>
</tr>
<tr>
<td>Ventral recumbency</td>
<td>Lying on abdomen/sternum with front legs folded under the body and visible hind legs (right side, left side); udder is partially obscured</td>
<td>[29]</td>
</tr>
</tbody>
</table>
Table 1. Cont.

<table>
<thead>
<tr>
<th>Posture Type</th>
<th>Description</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sternal recumbency/Lying on belly/Sternal lying/Lying on the stomach/Prone</td>
<td>Lying on abdomen/sternum with front and hind legs folded under the body; udder is totally obscured</td>
<td>[7,29,55–57]</td>
</tr>
<tr>
<td>Lateral recumbency/Lying on side/Lateral lying/Sidling</td>
<td>Lying on either side with all four legs visible (right side, left side); udder is totally visible</td>
<td>[7,29,55,57]</td>
</tr>
<tr>
<td>Sitting</td>
<td>Partly erected on stretched front legs with caudal end of body contacting the floor</td>
<td>[29]</td>
</tr>
<tr>
<td>Mounting</td>
<td>Stretch the hind legs and standing on the floor with the front legs in contact with the body of another pig</td>
<td>[55]</td>
</tr>
<tr>
<td>Exploring</td>
<td>The pig’s snout approached or nosed a part of the pen for more than 2 s. This was differentiated depending on the objects, such as a wall (including a fence door) or the floor</td>
<td>[42,58]</td>
</tr>
</tbody>
</table>

Before deep learning began to thrive in the field of machine vision, classifiers were commonly built on sophisticated handcrafted feature-extraction algorithms. Within them, the Zernike moment feature [59] is a common descriptor in early pig posture classification. The extracted features largely determine the classification performance of SVM, which has long been used for pig posture classification. In pig posture detection, due to the characteristics of translation, rotation, and scale invariance, the outline of pigs can serve as an effective and feasible feature representation. By inputting the Zernike moment feature extracted from the binary pig contours into an SVM classifier, four kinds of behavior postures of pigs can be recognized with a classification accuracy above 95% [40]. Similar to the above research, Nasirahmadi et al. [7] used the convex hull and boundaries of pigs as feature descriptors, obtaining a posture classification accuracy of 94.4% for two lying postures of grouped pigs.

Due to the background noise caused by occlusion and overlapping among pigs, posture classification based on manually designed feature descriptors lacks robustness in a real pig farming environment. Using a convolution operation coupled with deep learning, more semantic, high-level, and deeper representation of the features can be realized through a deep convolutional neural network (CNN) [54] in contrast to the shallow feature representation of the traditional method. With the help of the self-learning system, such as the gradient descent algorithm [60,61], the efficiency of feature extraction and classification can be greatly improved. Similar to the operation of Nasirahmadi et al., by introducing CNN to replace manual feature extraction, Xu [35] inputted the depth distance and boundary data of detected pigs in the depth image into a six-layer deep CNN. As a result, effective feature representation could be automatically and efficiently achieved from high-dimensional data. At the final stage, these scientific and powerful feature vectors can be fed into an SVM classifier, and five postures can be classified with an overall high accuracy of 94.6368%.

2.2.3. Pig Segmentation

Pig segmentation divides the pixels of an image into two subsets of foreground targets and background areas and generates pig segmentation masks [37], as Figure 5a shows. In traditional image processing, segmentation to obtain pig contours is used for localization prior to posture classification, which may affect the performance of pig classification since the quality of feature extraction of pigs is primarily determined by the effectiveness of pig segmentation [6]. In a previous study, by adopting the Otsu
adaptive threshold segmentation algorithm followed by canny edge detection and using morphological algorithms, normalized binary contour images of every pig were extracted for the classification of four kinds of behavioral postures of pigs [41]. Similar to threshold segmentation, watershed transformation can also separate the boundary of each pig from the background under commercial farm conditions for the detection of lateral and sternal lying postures [7].

In the deep learning field, segmentation technology always plays an assisting role rather than serves as an essential condition for pig posture detection; it can also accelerate the detection process and enhance the accuracy by providing rich information for subsequent processes under certain circumstances [62,63]. By integrating an advanced semantic-segmentation model DeepLabv3+ [64] into a pig posture detector, more contextual information can be obtained, which can greatly improve the feature extraction by the base network [42]. In a previous study [37], a panoptic segmentation combining semantic segmentation and instance segmentation can provide pig information with pixel-level accuracy for posture recognition, which can overcome the disadvantage of gross bounding boxes and sparse key points.

2.2.4. Pig Identification and Tracking

Pig identification and tracking refer to the recognition and preservation of the identity of each pig, such as a virtual unique identity or a given unique name, and simultaneous tracing of the trajectory of individual pigs throughout the whole monitoring period [65], as Figure 5b shows. Identification and tracking of individual pigs makes it possible to analyze the posture at the individual level rather than at the group level [6]. Through the integration of pig identification and tracking into posture detection, the abnormal behavior of a specific pig in a certain period can be estimated, which is significant for early risk alerts in pig farming [66].

However, for grouped pigs with highly similar appearances in intensive farming, it is challenging to apply machine vision technology. One suboptimal solution is to make color marks on the back of pigs, but it has little effect and cannot be sustained. With the assistance of non-visual methods, RFID tags with unique identifiers were permanently fixed to pig ears for the identification of pigs in pens [22]. However, this method is intrusive, vulnerable to disturbances, and requires significant labor to install. For higher feasibility, some researchers attempted to imitate human face recognition and adopted it to pig face recognition for discrimination of individual pigs [67,68]. However, the surveillance cameras used in practice are usually in a top view, which have difficulty capturing the entire face of pigs, and therefore pig face recognition is nearly impossible to apply in practice.

To solve this problem, the multiple object tracking (MOT) [69] technique was introduced. By integrating MOT algorithms, such as the Munkres variant of the Hungarian assignment algorithm and Kalman filters to the CNN-based pipeline (such as YOLOv2 [70] and Faster R-CNN), the pig detection model can preserve and recover the identity of pigs determined in the initial frame throughout the whole consecutive video sequence [28,71]. By integrating pig identification and tracking into posture detection, a pig profile comprising the posture, behavior, and position in a certain period can be generated for each individual in the study by Alameer et al. [28], which can provide valuable information for pig management.

3. Application of Deep Learning Methods to Pig Posture Detection

Many studies have demonstrated that the incorporation of deep learning techniques with image systems has superior performance and feasibility relative to traditional methods, which is expected to be the optimal solution to pig posture detection. As a typical implementation of deep learning methods, CNNs have become one of the most common models for various visual detection tasks and have been introduced into pig posture detection. Based on CNNs but with a different design concept, deep learning methods prevalent in pig posture detection can be mainly divided into two types: two-stage detection methods and
one-stage detection methods [72]. Two-stage detection methods usually have fairly high accuracy and robustness but a slow detection speed due to the need to generate a region proposal before the classification and localization of the object and refining the bounding boxes in the post-process stage [73]. A typical two-stage model generally consists of an RPN and a downstream detection network, as Figure 6 illustrates. Nevertheless, one-stage detection methods can directly estimate the classification accuracy and coordinate the position of objects in an image in only one evaluation without the need for region proposals, as Figure 7 demonstrates. A united network structure can confer one-stage detection methods with a significantly high detection speed when maintaining an acceptable accuracy [73].

3.1. Two-Stage Detection Models

R-CNN is a pioneer two-stage deep learning object detector, which is well-known for dramatically outperforming other contemporaneous state-of-the-art traditional machine learning detectors on the 200-class ILSVRC2013 detection dataset [74]. Based on R-CNN, Fast R-CNN and Faster R-CNN were successively proposed to solve the problems of computational redundancy and complex model composition, which could significantly improve the speed and accuracy. As the ultimate version of the R-CNN series, Faster R-CNN framework becomes the most popular among two-stage detection models to be used in pig posture detection [72]. To solve the problem of dependence on the facility environment of traditional machine vision systems, Riekert et al. [51] integrated the neural architecture search (NASNet) into the Faster R-CNN framework for pig localization and posture classification. As a result, by using only RGB images, they achieved a mean average precision (mAP) of 80.2% on pig position and posture detection with sufficient similar training images, but the mAP was only 44.8–58.8% under more difficult and demanding training conditions. To search for the optimal method for continuous posture monitoring of pigs throughout the day, Riekert et al. [75] tested 150 different deep learning model configurations in their experiments. Finally, a two-stage model consisting of Faster R-CNN and NASNet was identified as optimal despite certain flaws in speed. However, the method has comparatively poor performance at night due to the degraded near-infrared image quality.

Although Faster R-CNN greatly outperforms the traditional methods in terms of accuracy and speed, the detection speed is still inadequate to meet the demand of real-time tasks [46]. Hence, region-based fully convolutional networks (R-FCN) [76] and feature pyramid network (FPN) [77] were developed based on the two-stage concept to further improve the detection speed. R-FCN introduces the novel concept of position-sensitive RoI pooling and implements parameter sharing during calculation, which can significantly reduce the calculation time [76]. As a case in point, Nasirahmadi et al. [20] conducted a series of experiments to test three deep-learning-based detector frameworks (Faster R-CNN, Single Shot MultiBox Detector (SSD) [78], and R-FCN) using various feature extractors of RGB images separately. Finally, they demonstrated that R-FCN incorporating ResNet101 could achieve the highest mAP of 93% and an acceptable framerate, five frames per second, to detect standing and two different lying postures.

Because 2D images lack three-dimensional information and are susceptible to light disturbance, researchers began to explore the use of depth images. In a study to recognize and quantitively measure the posture of lactating sows during night and daytime, Zheng et al. [29] applied Faster R-CNN detectors to depth images to detect pigs and recognize their postures, including sitting and three different lying patterns, in loose pens. They found that the sows laid down more slowly to avoid crushing piglets. In this experiment, the use of depth images mitigated the disturbance of light during nighttime that may occur on RGB images. After the improvement to Faster R-CNN, Zhu et al. [79] detected five postures of lactating sows using the proposed refined two-stream RGB-D faster R-CNN model that used two CNNs to extract features from RGB images and depth images, respectively. The method they proposed takes advantage of feature concatenation fusion, contributing
to outstanding detection performance on the five postures, particularly for standing and lateral recumbency, which both have average recognition precision exceeding 99%.

The performance of two-stage methods for pig posture detection is summarized in Table 2. It can be seen that the methods using 2D images can barely detect the sitting posture and have difficulty distinguishing between sternal recumbency and ventral recumbency. However, with depth images, more posture types can be recognized, and the performance can be maintained at a competitive level relative to the use of 2D images.

**Table 2.** Summary of two-stage deep learning methods on pig posture detection.

<table>
<thead>
<tr>
<th>Posture Classification</th>
<th>Data Source</th>
<th>Detection Model</th>
<th>Evaluation Index</th>
<th>Results</th>
<th>Advantage</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lying and not lying</td>
<td>2D RGB image</td>
<td>Faster R-CNN + NASNet</td>
<td>AP(_{\text{localization}})@0.5</td>
<td>87.4%</td>
<td>Consider dataset diversity and focus on generalization under limited training data and different application settings</td>
<td>[51]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>mAP</td>
<td>80.2%</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>mAP@0.50 on day</td>
<td>84%</td>
<td>Enable continuous 24/7 pig posture detection and provide an optimal deep learning model after experimenting with over 150 different model configurations</td>
<td>[75]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Faster R-CNN + NASNet</td>
<td>mAP on day 84%</td>
<td>58%</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>mAP</td>
<td>89.9%</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Precision of posture classification</td>
<td>93%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standing, lying on side, and lying on belly</td>
<td>R-FCN + ResNet 101</td>
<td>mAP</td>
<td>&gt;93%</td>
<td></td>
<td>High flexibility and robustness</td>
<td>[20]</td>
</tr>
<tr>
<td>Standing, sitting, sternal recumbency, ventral recumbency, and lateral recumbency</td>
<td>Depth images</td>
<td>Faster R-CNN</td>
<td>mAP</td>
<td>87.1%</td>
<td>Reach real-time detection speed; Summarize the changing laws of pig position and posture</td>
<td>[29]</td>
</tr>
<tr>
<td></td>
<td>2D RGB image + Depth image</td>
<td>Refined two-stream RGB-D faster R-CNN</td>
<td>mAP</td>
<td>95.47%</td>
<td>Improve the accuracy of posture recognition by feature-level fusion strategy using RGB-D data</td>
<td>[79]</td>
</tr>
</tbody>
</table>

3.2. One-Stage Detection Models

YOLOv1 is the first one-stage model in the deep learning era, followed by subsequent versions (such as YOLOv2, YOLOv3 [80], and YOLOv5 [81]) and refined models based on them, have been increasingly developed and applied to pig posture detection, showing dramatic improvement of speed over two-stage detection models. In an experiment aimed at constructing high-quality pig posture datasets for the development of deep learning models, Kim et al. [60] found that, in a specific dataset, YOLOv2 could reach an average precision (AP) as high as 97%, which is seven times superior to that of SSD (a one-stage model). Alameer et al. [28] initiatively detected several pig postures at the individual level, including sitting, a tricky posture type, and additionally implemented pig identification and tracking without the use of any marks or sensors on pigs. In their research, two deep learning models were tested for behavior and posture detection, and they finally found that YOLOv2 was clearly superior to Faster R-CNN in terms of mAP and speed, with a high mAP above 98%. Furthermore, it can generate a profile for each pig that includes movement history and corresponding postures at a certain time point, which is valuable for commercial applications in automatic pig monitoring. Sivamani et al. [38] trained the tiny YOLOv3 model with datasets from nine pens to detect six pig postures. The results demonstrated that the model prevails over most two-stage deep learning models, such as Faster R-CNN, R-FCN, and machine learning models such as SVM [7], with a high mAP of 95.9%. Shao et al. [42] designed an assembled model for pig detection, segmentation, and classification, which are executed by YOLOv5, DeepLabv3+, and Resnet, respectively, and obtained a classification accuracy of 92.26% over four postures. Compared with static
posture detection, the method proposed by Shao et al. focuses on the detection of dynamic posture to facilitate detecting pathological changes in pigs.

Besides native YOLO frameworks, improved or refined models based on them have been designed for pig posture detection as well. By splitting pig posture detection into two separate modules, Witte and Marx Gómez [82] applied YOLOv5 for pig localization and EfficientNet-B0 for posture classification, achieving an AP of 99.4% and a precision of 93% for the two tasks, respectively. In terms of $\text{AP}^{\text{IoU}=0.5}$, precision, and recall, this approach greatly outperforms the use of YOLOv5 alone in the detection of lying and not-lying posture. Another improved model, improved YOLOX [56], was developed by incorporating effective tricks of YOLOv5 into YOLOX. Its performance was tested for detecting the standing, lying, and sitting postures of pigs, and it particularly exhibited improved performance in sitting recognition after simple data augmentation on sitting posture, with an AP0.5 of 90.9% and an AP0.5–0.95 of 82.8%. Compared with the most popular two-stage model Faster R-CNN, the improved YOLOX has about 10% higher mAP0.5–0.95, nearly 10-fold higher speed, and a much smaller model size, showing great superiority as a one-stage method. Huang et al. [57] designed a high-effect YOLO model that improved the performance of the feature extractor and integrated the dual attention mechanism. The high-effect YOLO model showed obvious improvement of mAP on the detection of standing, sitting, prone, and sidling posture compared with YOLOv3, SSD, and Faster R–CNN. Additionally, under occlusion and overlapping of pig bodies, the high-effect YOLO model displayed advantages in generalization and luminance robustness. Built on YOLOv5 architecture, a novel Light-SPD-YOLO model was developed by introducing an improved compressed block and a channel-wise mechanism, enhancing the speed and accuracy of the detection of five pig postures while lowering the model computation [55].

One-stage methods, as represented by YOLO series models, have exhibited powerful strength in detecting more complex postures on the basis of only using 2D images as shown in Table 3. Besides great superiority in speed, one-stage methods need only limited computer resources, but have comparable or even greater accuracy than two-stage methods, suggesting that one-stage methods may be more qualified for real application in commercial pig farms.

Table 3. Summary of one-stage deep learning methods on pig posture detection.

<table>
<thead>
<tr>
<th>Posture Classification</th>
<th>Data Source</th>
<th>Detection Model</th>
<th>Evaluation Index</th>
<th>Results</th>
<th>Advantage</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nine types of posture and behaviors</td>
<td>2D RGB image</td>
<td>YOLOv2</td>
<td>Accuracy</td>
<td>97%</td>
<td>Contribute high-quality datasets for building deep learning models; design the livestock safety surveillance system</td>
<td>[60]</td>
</tr>
<tr>
<td>Standing, lateral lying, sternal lying, and sitting</td>
<td></td>
<td>YOLOv2 + ResNet-50</td>
<td>mAP</td>
<td>98.88%</td>
<td>Enable robust and accurate monitoring of individual pigs; generate a profile of each pig</td>
<td>[28]</td>
</tr>
<tr>
<td>Sitting, lying, Standing, multi, part-of, and other</td>
<td></td>
<td>Tiny YOLOv3</td>
<td>mAP</td>
<td>95.9%</td>
<td>High accuracy with fewer computational resources</td>
<td>[38]</td>
</tr>
<tr>
<td>Standing, lying on the stomach, lying on the side, and exploring</td>
<td></td>
<td>YOLOv5 + DeepLabv3+ + Resnet</td>
<td>Classification accuracy</td>
<td>92.26%</td>
<td>Propose a joint training method that involves pig posture detection and pig semantic segmentation</td>
<td>[42]</td>
</tr>
</tbody>
</table>
### Table 3. Cont.

<table>
<thead>
<tr>
<th>Posture Classification</th>
<th>Data Source</th>
<th>Detection Model</th>
<th>Evaluation Index</th>
<th>Results</th>
<th>Advantage</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standing, lying, and sitting</td>
<td>Improved YOLOX</td>
<td>AP&lt;sub&gt;0.5&lt;/sub&gt; for localization</td>
<td>99.5%</td>
<td></td>
<td>Focus on sitting detection; solve class imbalance on the lack of sitting posture</td>
<td>[56]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AP&lt;sub&gt;0.5-0.95&lt;/sub&gt; for localization</td>
<td>91%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>mAP&lt;sub&gt;0.5&lt;/sub&gt;</td>
<td>95.7%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>mAP&lt;sub&gt;0.5-0.95&lt;/sub&gt;</td>
<td>87.2%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standing, sitting, prone, and sidling</td>
<td>High-effect YOLO</td>
<td>mAP</td>
<td>97.43%</td>
<td></td>
<td>Optimize the YOLO v3 model; construct multiple attention mechanism</td>
<td>[57]</td>
</tr>
<tr>
<td>Standing, lying on the belly, lying on the side, sitting, and mounting</td>
<td>Light-SPD-YOLO</td>
<td>mAP</td>
<td>92.04%</td>
<td></td>
<td>High speed and accuracy with low model parameters and the computation complexity</td>
<td>[55]</td>
</tr>
<tr>
<td>Lying and not lying</td>
<td>YOLOv5 + EfficientNet</td>
<td>AP&lt;sub&gt;IoU=0.5&lt;/sub&gt; of localization</td>
<td>99.4%</td>
<td></td>
<td>Achieve significant improvement in accuracy by considering pig posture classification as a two-step classification process</td>
<td>[82]</td>
</tr>
</tbody>
</table>

### 3.3. Other Deep Learning Detection Methods

Segmentation and extraction of pig contours from binary images by segmentation algorithms (such as threshold segmentation algorithm and watershed transformation) were once the dominant approaches for detecting pigs in 2D images or depth images. On this basis, the incorporation of deep learning technologies can greatly improve performance. Xu et al. [35] replaced traditional feature extraction methods with a CNN to deeply extract features from various parameters related to depth distance and pig contours. Their method could achieve a classification accuracy of 94.63% over five posture types, which outperforms earlier methods relying on manual feature extraction. The study of Brunger et al. [37] is another example of the efficacy of deep learning in pig contour extraction. In their work, binary segmentation was first executed by a neural network instead of traditional threshold segmentation algorithms, and then the panoptic segmentation was obtained by combining instance segmentation. This compound segmentation method enables the extraction of individual pigs at pixel-level accuracy, thereby providing valuable information for future pig posture recognition. Ocepek et al. [83] used Mask R-CNN [84] to segment the body of pigs and determine whether their postures were curved or straight. In addition, a YOLOv4 [85] model was used to improve the detection of tails, with an average precision of about 90% to serve as an alternative to Mask R-CNN.

In contrast to common anchor-based techniques, such as typical two-stage and one-stage models, anchor-free pig posture detection models have the great ability to independently learn the anchor scale and thereby obviate the necessity of pre-establishment of anchors, which can notably reduce the computation complexity [52]. Moreover, Psota et al. [40] have developed a method that omits the use of bounding boxes and directly detects the precise posture of animals through key points located on specific body parts, including the shoulder and back. Gan et al. [86] designed a CNN-based key point detector, which excels at identifying two key points in piglets, namely the snout and hip. In the study of Wang et al. [39], an anchor-free CenterNet model was deployed to discriminate the standing posture.
and lying posture with an AP of 86.5%, and then the HRNet with Swin Transformer block (HRST) was designed for detecting the joint points of pigs in standing posture, which reached 40 frames per second in speed.

4. Discussion

In this section, to better assist incoming researchers in understanding the current state of the methods being advanced for pig posture detection, the literature discussed was primarily published after 2020, which conforms to the criteria of representing up-to-date technological progresses, exposing common limitations in existing research that remain unresolved, or exploring future directions and novel approaches. The limitations of present pig posture detection are outlined in Table 4 as a quick overview. For each identified limitation, specialized approaches from several disparate aspects are proposed as potential solutions. Subsequently, a brief blueprint of the development directions for pig posture detection methods utilizing deep learning will be presented.

<table>
<thead>
<tr>
<th>Limitation</th>
<th>Proposed Solutions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lack of standardized and comprehensive databases.</td>
<td>Collaboration for database construction and open-source data initiatives</td>
</tr>
<tr>
<td>Class imbalance.</td>
<td>Employment of data augmentation, transfer learning, and ensemble models</td>
</tr>
<tr>
<td>Demerit of single-view camera systems.</td>
<td>Adoption of depth cameras or multi-modal models compatible with different views</td>
</tr>
<tr>
<td>The challenge in multi-pig identification and tracking.</td>
<td>Combination with MOT technique</td>
</tr>
<tr>
<td>Neglect of temporal information.</td>
<td>Introduction of video-based detection and spatial-temporal model</td>
</tr>
</tbody>
</table>

4.1. Limitations of Current Methods and Viable Solutions

In the field of AI, training a model requires an enormous number of annotated datasets, such as ImageNet [74]. Similarly, for the convenience of further research carried out by subsequent researchers, large-scale pig datasets have been sequentially developed and released by preceding researchers [42,51,56,60,75]. Nevertheless, there currently is no standardized and comprehensive high-quality database available that covers all types of pig postures in a wide range of label formats. Therefore, collaboration between researchers and institutions, as well as open-source data initiatives, are encouraged to construct such a pig posture database for superior consistency and interoperability among different datasets. Another limitation in pig posture datasets is class imbalance that is generally caused by the excess (such as lying posture) or scarcity (such as sitting and mounting posture) of certain classes, which is not under human control, resulting in low detection accuracy of these specific classes [51,55,56]. Generating synthetic samples or variations of underrepresented classes, such as flipping, rotating, scaling, or introducing noise, to balance the dataset is one way to approach the problem, but it has a limited effect [56]. From the aspect of models, the transfer learning technique cooperating with ensemble models offers a promising solution to the issue by utilizing pre-trained models or knowledge from related tasks and combining the predictions of multiple models [87]. One of the attempts to tackle the limitation in the pig posture detection domain is to endow specific classes with different weights during training [82], such as weighing minority classes more than other classes.

Current detection methods under development primarily rely on a single camera, meaning that only a singular angle of pigs can be observed [42,56]. Under a single-view system, the phenomenon of adhesion and overlapping among captive pigs in the monitoring image hinders the ability to detect pig postures correctly [57]. Additionally, the differentiation between sternal recumbency and ventral recumbency is still a challenge for
A viable solution is to substitute depth cameras for 2D cameras or to add 2D cameras aimed in other directions. In actual farm environments, to expand the scope of pig monitoring, multiple cameras with different views have been used. However, existing models specifically designed for single-view posture detection are not compatible with multi-view scenarios due to inadequate learning ability [56]. To solve the dilemma, on the one hand, researchers may design and develop new model architectures by incorporating the multi-modal fusion technique [88] to concurrently process images from different devices. On the other hand, the construction of multi-view pig posture datasets is required for long-term progress.

Moreover, identification and tracking of multiple pigs throughout the real-time monitor video are neglected by most of the existing research, and it also has certain challenges due to interactions and occlusions frequently appearing in crowded environments as well as the inherent nature of the similarity in the appearance of pigs. MOT techniques, such as the Byte Track algorithm [89], can be introduced to address continuous tracking and new detection ideas, such as the anchor-free approach [90], may be helpful to deal with crowded scenarios. Additionally, it leads to another limitation that currently proposed models are primarily designed to detect and analyze postures in static images or in discrete frames extracted from a video sequence, which may result in the neglect of models for temporal information about pig movement and posture change between contiguous frames. Instead, video-based detection can learn the patterns and transitions of pig postures over time [72], enhancing the comprehensive understanding of pig behaviors. Models that can maintain a memory of previous spatial-temporal feature information, such as the recurrent neural network (RNN) [91] and the temporal convolutional networks (TCNs) [92], are promising approaches to the problem. The work of Zhang is an instance that utilizes an inflated 3D convolutional network to extract temporal and spatial information of pigs’ behaviors from image frames and stacked optical flow, respectively [93]. In addition, corresponding large datasets of labeled video footage will be required to train spatial-temporal models.

### 4.2. Outlook of Pig Posture Detection Methods with Deep Learning

Promising development and latent directions in pig posture detection methods can focus on the improvement of the algorithm and model themselves, full utilization of diverse data, and the connection with real-world issues to address current challenges and advance the feasibility for the broader application. Firstly, high accuracy and speed remain to be crucial factors to be considered in the future. Consideration of new deep learning optimization techniques, such as the genetic algorithm-based cuckoo search algorithm that has been proven to be highly effective in the classification of fish [94], may be needed in the future. Then, the model’s computation complexity and resource demand should be reduced, making it applicable for low-computation-power platforms, such as mobile devices. Additionally, future research should focus on the identification and tracking of multiple pigs throughout monitoring while accurately detecting the posture of each pig, and occlusion handling must be considered for tracking continuity and temporal information analysis to allow deeper insights into the postures each pig shows. Furthermore, to support the advancements of research within or even beyond this field, a large-scale, standardized, and comprehensive database should be constructed. On the one hand, unified datasets can conduct to the validation and benchmarking for evaluating and comparing different deep learning models. On the other hand, an all-inclusive database should encompass diverse attributes related to pigs, such as pig breeds, pig physical condition, pig movement history, and breeding conditions, in addition to the pig images or videos. To leverage these different types of information, multi-modal learning technology should be introduced and adopted to current deep learning models to strengthen the evaluation results.

### 5. Conclusions

Numerous studies have proven that pig postures under different experimental conditions can reflect the health and welfare of pigs and predict abnormal events. Accurate and
powerful automatic posture detection methods are the basis for posture analysis. Initially, sensors were used as data acquisition equipment, but they cannot meet the demand for contact-free, stress-free, and automatic operations. Then, with the advancement of image processing technologies, cameras were widely used in the design and development of automatic detection methods. Depth cameras can obtain depth information in addition to 2D data in each pixel area to generate more convictive detection results. However, 2D cameras are much more popular in pig posture detection for their low cost and convenience, and the lack of three-dimensional information can be solved by installing multiple cameras aimed in different directions. In terms of posture detection methods, the traditional machine vision pipeline generally consists of pig region selection, feature extraction, and final classification. What distinguishes deep learning methods from traditional methods is that the manually designed algorithms in these processes can be replaced by self-learning systems, particularly for feature extraction. This remarkable property is mainly ascribed to the application of deep neural networks, or more precisely, deep CNN. In the posture detection field, the Faster R-CNN and YOLO series models are the most popular deep learning models, which have advantages in detection accuracy and detection speed, respectively. In recent years, methods based on the anchor-free principle have successfully achieved the trade-off between accuracy and speed. Among various models proposed so far, some are powerful enough in certain aspects of the practical demand of pig farming.

In conclusion, as a dominant developing trend, deep learning methods incorporated with image systems have achieved great success in research on pig posture detection. Although existing methods expose certain limitations on several aspects, they still show promising potential to be popularized in the commercial pig farming industry on a large scale.

Author Contributions: Conceptualization, Z.C. and H.W.; methodology, Z.C. and J.L.; investigation, Z.C.; resources, H.W.; data curation, J.L.; writing—original draft preparation, Z.C.; writing—review and editing, Z.C., J.L. and H.W.; supervision, H.W.; funding acquisition, H.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Fundamental Research Funds for the Central Universities of China (2662022XXYJ009), the National key research and development program (2022YFD1601903), the Hubei Province Science and Technology Major Project (2022ABA002), and the HZAU-AGIS Cooperation Fund (SZYJY2022034).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Conflicts of Interest: The authors declare no conflict of interest.

References
1. Iglesias, P.M.; Camerlink, I. Tail posture and motion in relation to natural behaviour in juvenile and adult pigs. *Animal* 2022, 16, 100489. [CrossRef] [PubMed]


Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.