Skeleton extraction and pose estimation of piglets using ZS-DLC-PAF

Chengqi Liu[1], Haijian Ye[1], Shuhan Lu[2], Zhan Tang[1], Zhao Bai[1], Lei Diao[1], Longhe Wang[3], Lin Li[1,2*]

(1. *College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, China*;
2. *School of Information, University of Michigan, Ann Arbor 48109, USA*;
3. *National Research Facility for Phenotypic and Genotypic Analysis of Model Animals (Beijing), Beijing 100083, China*)

Abstract: The accurate identification of various postures in the daily life of piglets that are directly reflected by their skeleton morphology is necessary to study the behavioral characteristics of pigs. Accordingly, this study proposed a novel approach for the skeleton extraction and pose estimation of piglets. First, an improved Zhang-Suen (ZS) thinning algorithm based on morphology was used to establish the chain code mechanism of the burr and the redundant information deletion templates to achieve a single-pixel width extraction of pig skeletons. Then, body nodes were extracted on the basis of the improved DeepLabCut (DLC) algorithm, and a part affinity field (PAF) was added to realize the connection of body nodes, and consequently, construct a database of pig behavior and postures. Finally, a support vector machine was used for pose matching to recognize the main behavior of piglets. In this study, 14 000 images of piglets with different types of behavior were used in posture recognition experiments. Results showed that the improved algorithm based on ZS-DLC-PAF achieved the best thinning rate compared with those of distance transformation, medial axis transformation, morphology refinement, and the traditional ZS algorithm. The node tracking accuracy reached 85.08%, and the pressure test could accurately detect up to 35 nodes of 5 pigs. The average accuracy of posture matching was 89.60%. This study not only realized the single-pixel extraction of piglets' skeletons but also the connection among the different behavior body nodes of individual sows and multiple piglets. Furthermore, this study established a database of pig posture behavior, which provides a reference for studying animal behavior identification and classification and anomaly detection.

Keywords: piglets, skeleton extraction, pose estimation, Zhang-Suen, DeepLabCut, Part affinity field
DOI: 10.25165/j.ijabe.20231603.6930

Citation: Liu C Q, Ye H J, Lu S H, Tang Z, Bai Z, Diao L, et al. Skeleton extraction and pose estimation of piglets using ZS-DLC-PAF. Int J Agric & Biol Eng, 2023; 16(3): 180–193.

1 Introduction

The accurate identification of various postures of pigs in their daily life is crucial for studying the behavioral characteristics of pigs. In the field of pose recognition, skeleton parameters are typically used to represent the difference between pose and action, such as lifting the head and crouching[1]. Skeleton extraction refers to the extraction of the linear skeleton of an object through iterative shrinkage and other methods; the extracted skeleton is then used to directly reflect the structural features and geometric shapes of an object[2]. Existing pose recognition algorithms are mostly suitable for human research[3], and no skeleton posture database is built specifically for pigs. However, traditional thinning algorithms are highly sensitive to the noise of a pig's target boundary in a video considering that the environment in a piggery is dirty and dark. Traditional algorithms exhibit the disadvantage of repeatedly calculating the reservation point, and there are a lot of burrs and redundant information[4]. Therefore, finding a novel approach for pig behavior skeleton extraction is necessary to identify pig posture accurately. Then, an effective pig posture database can be established to support pig behavior research[5-7].

Image skeleton extraction methods are mostly based on geometry and learning[8]. Geometry-based methods include morphology refinement (MR)[9], distance transformation (DT)[10], and medial axis transformation (MAT)[11]. Among them, MR exhibits the advantage of maintaining skeleton continuity; however, its result will deviate from the real skeleton. DT can quickly extract the skeleton, but it cannot guarantee the skeleton's continuity. MAT is good at maintaining the topological characteristics of the original image, but it is slower and requires more iterations. Moreover, it is unable to achieve single-pixel skeleton extraction in the diagonal part. Lynda et al.[12] proposed the modified Zhang-Suen (MZS) algorithm that can completely retain the 2×2 square structure. However, the parity of the pixel coordinate index values will directly affect the results. To increase speed, a general parallel computing platform and programming model were introduced to accelerate the algorithm[13]. Li et al.[14] pointed out that an uneven surface contour will lead to a problem in the thinning results, i.e., the bifurcation outline of the image will be affected by noise. The skeleton extracted using the aforementioned methods still suffers from confusion of priorities and the problem of a disordered structure. Improving existing methods to enable them to judge the shape and connectivity of the original object accurately is the key to constructing a skeleton posture database in the current study.

Received date: 2021-07-26 **Accepted date:** 2022-06-05
Biographies: Chengqi Liu, PhD candidate, research interest: image processing, Email: liuchengqi@tsinghua.edu.cn; **Haijian Ye**, Master, Professor, research interest: image processing and big data, Email: hjye@cau.edu.cn; **Shuhan Lu**, Master candidate, research interest: health information, Email: shuhanlu@umich.edu; **Zhan Tang**, PhD candidate, research interest: text mining, Email: tz_blues@163.com; **Zhao Bai**, Master candidate, research interest: image processing, Email: s20193081367@cau.edu.cn; **Lei Diao**, Master candidate, research interest: text mining, Email: S20193081368@cau.edu.cn; **Longhe Wang**, Master candidate, research interest: data mining, Email: phil.wang@cau.edu.cn.

***Corresponding author: Lin Li**, PhD, Professor, research interest: image processing and text mining. College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, China. Tel: +86-13621097034, Email: lilincau@126.com.

On the basis of a deep learning method, posture features have been extracted by designing a deep learning model to extract the skeleton accurately. Pfister et al.[15] extracted the internal relations between nodes through a spatial fusion model and then used the optical flow information to predict the heat map of adjacent frames; however, this method cannot output nodes for the entire body. Newell et al.[16] detected nodes and matched them by groups; the detection effect was better than those of the mask R-CNN and Google detection methods. Fang et al.[17] identified a segmented single posture and overcame the positioning error caused by detection. Liao et al.[18] used OpenPose to extract nodes and completed cross-perspective recognition on the basis of a pose-based temporal-spatial network and 3D pose estimation. With Faster R-CNN as a pig detector, the Kalman Filter has been adopted in big white pig tracking by Cowton[19]. Gan et al.[20] developed an online piglet tracking network (OPTN) composed of a base network, a detection head, and an association head. However, there is no in-depth study on the tracking of different nodes of piglets. The aforementioned deep learning methods mostly establish a human node model to detect posture, but the accuracy of node extraction for pigs remains to be investigated. In addition, the problems of long training time and relying on manual annotation persist. Gan et al.[21] continued to develop a deep learning network system mainly composed of three convolutional neural networks: a piglet detection head, a key point detector, and a piglet association head. In their study, piglet body key points (Snout and Hip) were detected precisely by using a CNN key point detector. Then, Gan et al.[22] continued to propose a fast and accurate detection method for sow nursing behavior using CNN-based optical flow and features. Both of them have achieved relatively high accuracy and achieved partial detection of key points, but there are also false detection cases affected by illumination. The DeepLabCut (DLC) algorithm[23] has been used for the pose estimation of experimental animals (e.g., mice and fruit flies) in high-definition videos by utilizing its advantages of providing a robust model and small sample labeling[24-26]. The multi-target tracking accuracy of DLC can reach more than 95%, and running efficiency is also improved. DLC provides the basis for realizing pig posture analysis in the current study. Therefore, another key research issue is how to apply the DLC method to obtain an accurate connection of pig nodes, enhancing the robustness of the models for various factors, such as specific pig parts and cross occlusion[27-29].

In summary, the methods for the skeleton extraction and pose estimation of piglets exhibit the following problems:

1) The accuracy of skeleton extraction must be improved, and skeleton problems of breakpoints, burr, and redundancy must be perfected;

2) No standardized behavior posture database for pigs for realizing behavior recognition and classification is available at present;

3) The connection of multi-objective similar joints and the distinction of different joints in the group of pigs have not yet been solved.

To address the aforementioned problems, the current study proposed a skeleton extraction and posture estimation method based on the ZS-DLC-part affinity field (PAF). The method presented in the current study realized the recognition of different types of pig behavior from the perspective of skeleton nodes. It recognized different poses with good robustness by using skeleton information, providing a new method that supports animal pose recognition.

2 Materials and methods

On the basis of the ZS-DLC-PAF algorithm, the current study realized the precise skeleton extraction and pose recognition of pigs. Its technical route was as follows:

1) Using a morphological thinning method, the skeleton was extracted from the preprocessed binary image of a pig target, and a skeleton extraction model was established;

2) On the basis of an improved ZS burr removal algorithm, the standard skeletons of pigs with different types of behavior without the burr effect were extracted, and a modified burr model was established;

3) On the basis of the human skeleton model, it optimized the main nodes of pigs, established posture evaluation criteria, and construct a database of pig behavior posture;

4) An improved DLC neural network model was adopted to extract the nodes associated with piglets' behavior;

5) PAF was introduced to connect individual nodes and complete posture database matching.

The specific steps of the algorithm are illustrated in Figure 1.

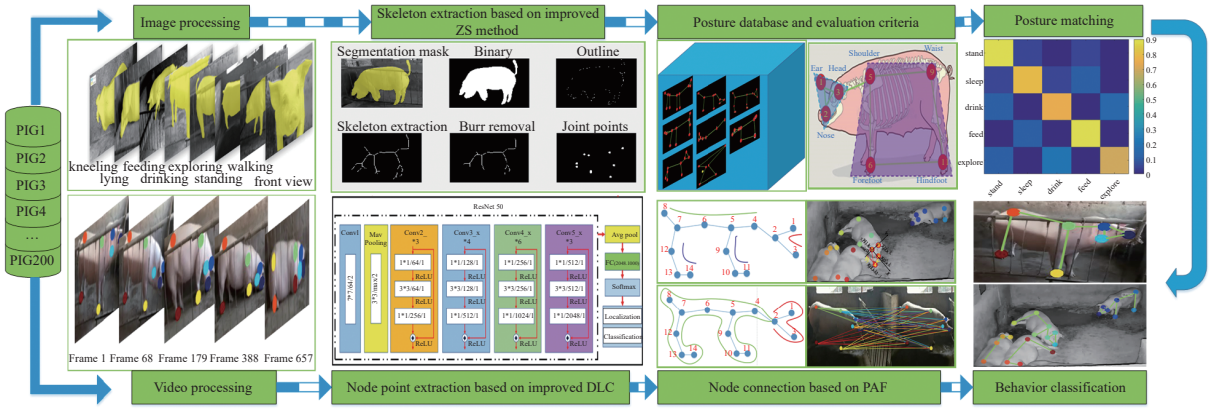


Figure 1 Steps of the method proposed in the current study

2.1 Subject and environment

2.1.1 Data collection

Experimental data were collected from a large-scale pig farm in Qingdao, and the subjects were individual sows and piglets in different stages of growth. The surveillance videos captured by a Hikvision smart ball camera (DS-2DE4320IW-DS-2DE4320IW-D) installed in the piggery were used as the experimental data. In the current study, 5000 videos of pigs were collected, 14 000 images with different types of behavior were selected as the training set, 2000 frames were selected as the verification set, and 2000 frames

were selected as the test set. The test set had different interference factors, such as occlusion, darkness, and blur, for analyzing the robustness of a model. Sample information is listed in Table 1. The description of different behavior characteristics is listed in Table 2.

Table 1 Statistics of pig images

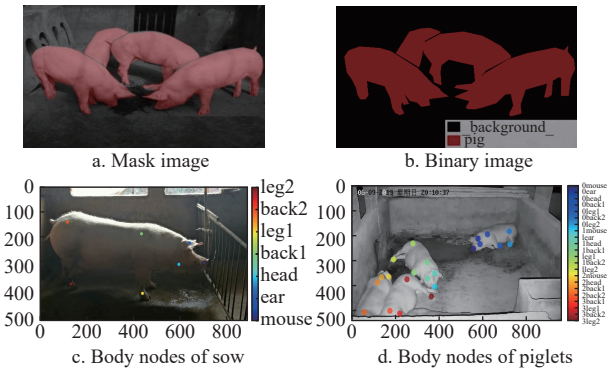
Data	Frames	Details	Body nodes
Training set	14 000	single-objective	1-7
Verification set	2000	single-objective	1-7
Test set	2000	single- and multi-objective	20-35

Table 2 Descriptions of pig behavior

Behavior	Description
Kneeling	The pig kneels down on its front and hind legs, and its body covers all its limbs in a straight line.
Lying	The body is flat on one side, with the two front legs and two hind legs basically overlapping.
Standing	With all four limbs on the ground, the body is upright and motionless, and the head is basically flat with the body.
Feeding	The pig leans forward, lowers its head, and keeps its legs still.
Exploring	The head is lowered, the mouth arches the ground, and the limbs are stepping on the ground.
Drinking	The head is raised, the snout is in contact with the drinking fountain, and all four limbs are still on the ground.
Walking	The body of the pig is upright, and its limbs rise and fall alternately with visible angles.
Watching	The pig's eyes are open. This pose is used as a reference for sleeping behavior.

2.1.2 Data Labeling

The annotated information used in the current study included position, contour, and skeleton information connected by seven key points in a specified order. A labeling tool, i.e., LabelMe, was used to segment a pig's body regions and calibrate nodes. As shown in Figure 2, polygons were used to mark the contour of the pigs in the current study, and the inflection point was as close as possible to the actual contour when marking. Given that a pig's neck, back, tail, and symmetrical parts exerted minimal influence on estimating its posture, the parts with higher recognition, including the ear, head, nose, shoulder, waist, forefoot, and hindfoot, were selected as the primary nodes (No. 1-7).



a. Mask image b. Binary image

c. Body nodes of sow d. Body nodes of piglets

Figure 2 Data labeling example of this study

2.2 Skeleton Extraction Algorithm

To solve the problems of redundancy and refinement of pig skeleton information in a dirty and dark pigsty, this study adopted an improved ZS skeleton extraction algorithm based on morphology.

2.2.1 Basic Definition

For pixel point P0, the pixel point sets in 8 and 24 neighborhoods are shown in Figure 3. P0 was the scan point, and P1-P8 corresponded to the chain code[30] in directions 1-8. The pixel point set P9-P24 was expanded by a layer around P1-P8 to obtain the pixels in 24 neighborhoods. The end, node, and growth structures of the skeleton were also set as shown in Figure 3.

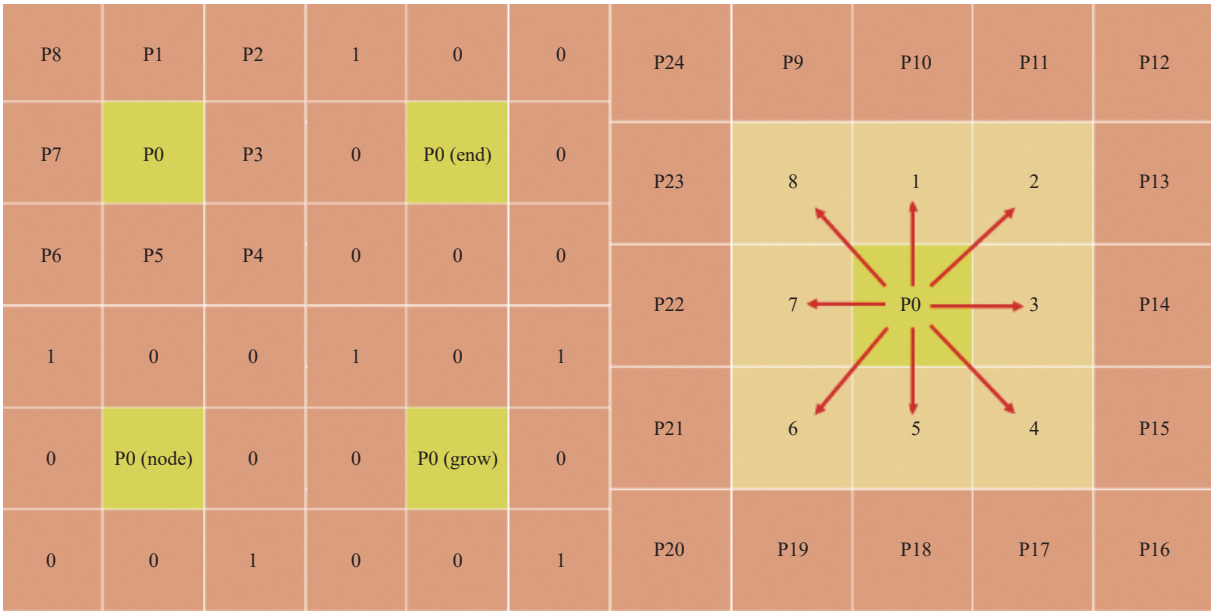


Figure 3 8 and 24 neighborhoods of point P0

The crossover number of P0, i.e., A(P0), was defined as the number of times that a pixel changed from 1 to 0 by circling clockwise along its eight neighborhoods. The number of foreground pixels in its eight neighborhoods was defined as B(P0). The number of foreground pixels in its 24 neighborhoods was defined as M(P0).

$$B(\mathrm{P0}) = \sum_{i=1}^{8} P_i \tag{1}$$

$$M(\mathrm{P0}) = \sum_{i=9}^{24} P_i \tag{2}$$

$$A(P0) = \left(|P_8 - P_1| + \sum_{i=1}^{8} |P_i - P_{i-1}| \right) / 2 \tag{3}$$

End: If only one skeleton point exists in the eight neighborhoods of P0 and P0 is a skeleton point, then P0 is called the endpoint.

$$P_{\text{end}} = \begin{cases} \text{true(count = 1)} \\ \text{false} \end{cases} \tag{4}$$

where, count is the total number of skeleton points in the eight neighborhoods of P0.

Node: If two skeleton points exist in the eight neighborhoods of P0, then P0 is called a node point.

$$P_{\text{node}} = \begin{cases} \text{true(count = 2)} \\ \text{false} \end{cases} \tag{5}$$

Grow: If three or more skeleton points exist in the eight neighborhoods of P0 and they are the starting points of a burr, then P0 is called the growing point.

$$P_{\text{grow}} = \begin{cases} \text{true(change = 3)} \\ \text{false} \end{cases} \tag{6}$$

where, change indicates the number of changes from a skeleton point to a background point in the eight neighborhoods of P0.

Burr: A burr is a branch in the skeleton that does not reflect structural information about the target. Combined with the number of iterations, the judging threshold of a burr is

$$L = \text{ceil}(2\sqrt{2} \times \text{times}) \tag{7}$$

where, L is the burr length, ceil indicates taking the smallest integer greater than or equal to the parentheses, and times denotes the number of iterations of image thinning.

Step size M: The number of all pixel points in a skeleton branch with a single pixel width.

2.2.2 Improved ZS Method

To reduce the redundant information and burr of the skeleton, the ZS algorithm was first adopted to delete the marks of non-skeleton points. The objectives of this algorithm were as follows:

1) To determine whether P0 was an endpoint. If P0 had only one adjacent point, then it was an endpoint and could not be marked. If P0 had seven adjacent points, then they could not be marked to ensure skeleton connectivity;

2) To detect whether a change occurred between 0 and 1 in the eight neighborhoods of P0 to ensure that the skeleton points were not marked.

3) To mark the non-skeleton points in the southeast side and northwest corner of the eight neighborhoods and then delete them;

4) Repeat (1) and (2), and then mark the non-skeleton points in the northwest side and southeast corner of the eight neighborhoods and delete them.

$$2 \le B(\text{P0}) \le 6 \tag{8}$$

$$A(\text{P0}) = 1 \tag{9}$$

$$\text{P1} \times \text{P3} \times \text{P5} = 0 \tag{10}$$

$$\text{P3} \times \text{P5} \times \text{P7} = 0 \tag{11}$$

$$\text{P1} \times \text{P5} \times \text{P7} = 0 \tag{12}$$

$$\text{P1} \times \text{P3} \times \text{P7} = 0 \tag{13}$$

The initial skeleton information of pig behavior was obtained through the preceding equations. Problems, such as the redundancy of growth points, an obstinate burr in the oblique direction, and the non-single pixel width of the skeleton, which cannot clearly describe behavior posture, still existed, as shown in Figure 4. Therefore, the main parts of algorithm improvement in the current study were as follows: the addition of a redundant elimination module and the establishment of the chain code mechanism of burr deletion.

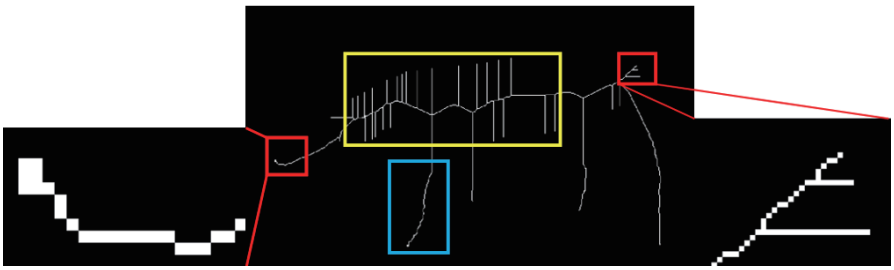


Figure 4 Skeleton burr and redundant information

The standard skeleton of a pig consists of seven endpoints: the ear end, tail end, mouth end, two hind limbs, and two forelimbs. The directional chain code was used to traverse the skeleton branch. For an effective branch, the thresholds of step size M and burr length L were set to delete the branch burr and return to the node in time. The ideal skeleton image and traversal order of skeleton nodes obtained after thinning are shown in Figure 5.

In accordance with the origin point $O(x_0, y_0)$ and the end coordinate (x_i, y_i), the angle θ between each endpoint and the center of mass was calculated to construct the evaluation standard of a posture database. The calculation formula for angle θ in each quadrant is as follows:

$$\theta = \arctan\left(\frac{|y - y0|}{|x - x0|}\right) + \frac{i\pi}{2}(i = 0, 1, 2, 3) \tag{14}$$

The primary reason for the non-single pixel width of the skeleton was that the points did not meet $B(P0)$=1. These points were not marked and deleted. Therefore, the redundant pixel deletion templates were designed as shown in Figure 6.

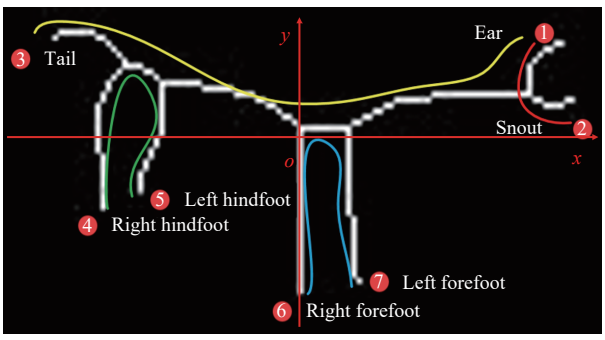


Figure 5 Ideal skeleton image and traversal order of skeleton nodes

0	1	X	X	0	0	X	1	0	0	0	X	0	1	0
1	P0	0	1	P0	0	0	P0	1	0	P0	1	1	P0	1
X	0	0	0	1	X	0	0	X	X	1	0	0	0	0

a. (P1 = P7 = 1) && (P3 = P4 = P5 = P8 = 0) b. (P5 = P7 = 1) && (P1 = P2 = P3 = P6 = 0) c. (P1 = P3 = 1) && (P2 = P5 = P6 = P7 = 0) d. (P1 = P4 = P7 = P8 = 0) && (P1 = P4 = P7 = P8 = 0) e. P2 + P4 + P6 + P8 = 0 && P1 + P3 + P5 + P7 = 3

Figure 6 Design of the redundant pixel deletion templates

2.3 Construction of behavior posture database

2.3.1 Body nodes optimization

In this study, the initial skeleton nodes of pigs were designed by referring to the skeleton nodes of the human body and then combining them with the characteristics of a pig skeleton, as shown in Figures 7 and 8. More than a dozen nodes were eventually retained due to the number of frontal limb nodes in the human body. The behavior posture of a pig mostly focuses on the angle of its head and the movement of its legs. When standing still, regardless of whether it is drinking, exploring, or eating, a pig is judged in accordance with the up and down movements of its head. When lying down, the points of the back and feet were judged as the side-lying or kneeling position. In addition, considering the different nodes extracted from the front, side, and aerial shot angles, 12 nodes of the head, limb, and back with common representatives were selected in the current study to form the pose model, which primarily included the left ear, right ear, nose, head, neck, shoulder, back, waist, tail, left forefoot, right forefoot, left hindfoot, and right hind foot.

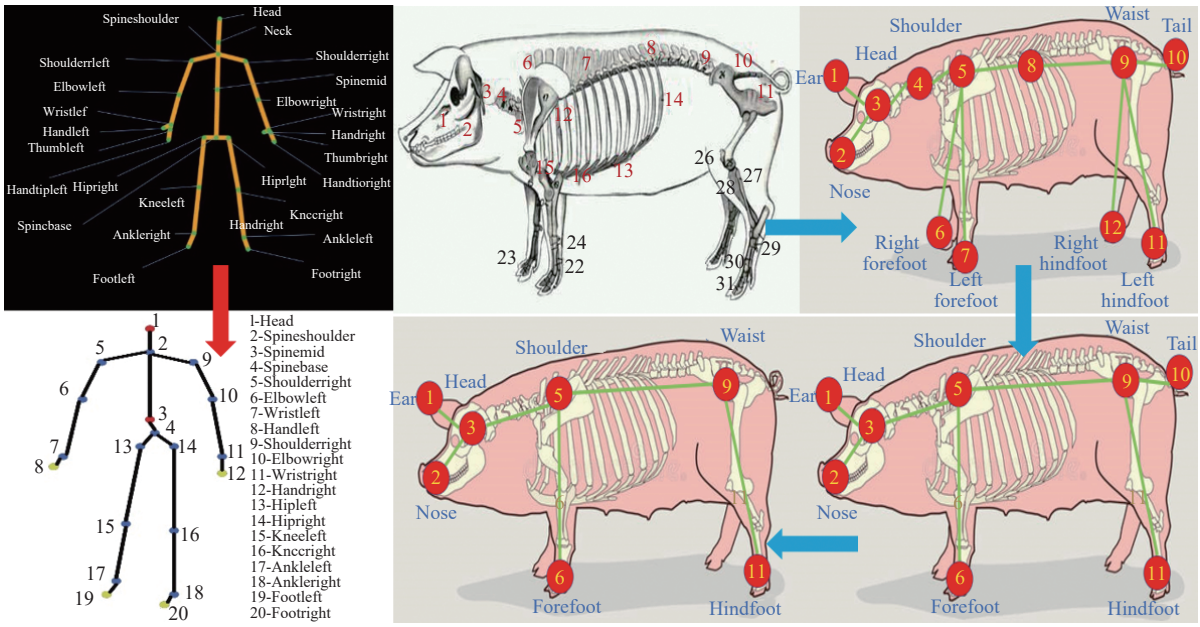


Figure 7 Design of the initial skeleton nodes of the overall pig

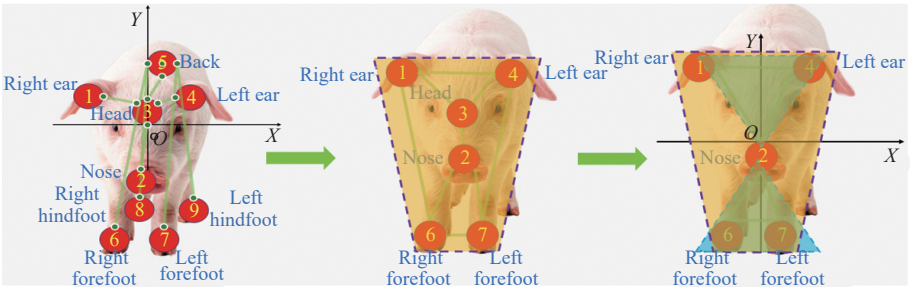
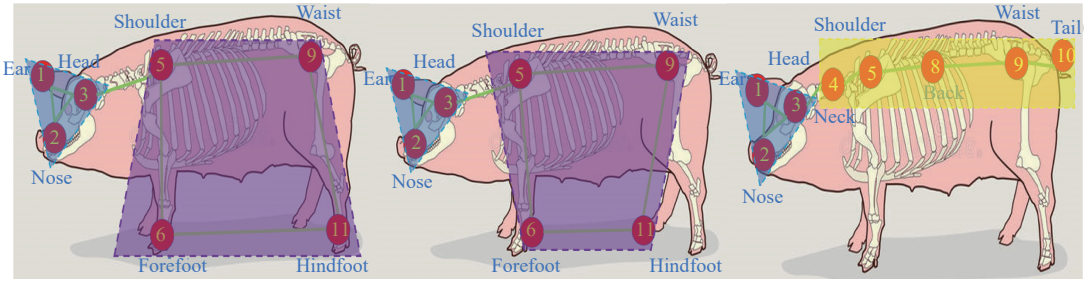


Figure 8 Design of the initial skeleton nodes of the head of a pig

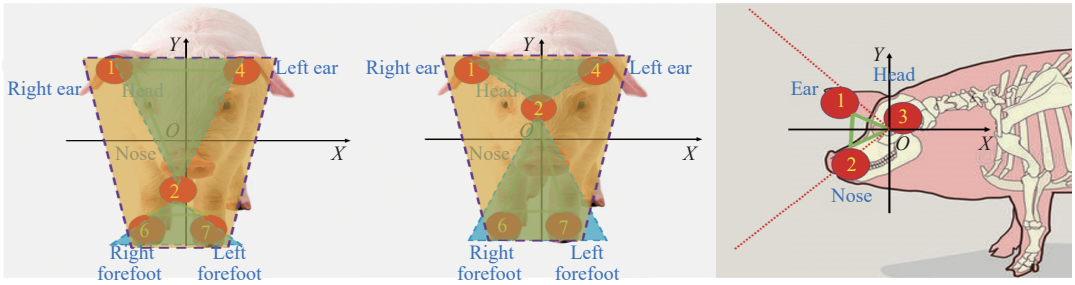
2.3.2 Postural evaluation criteria

In the current study, the head–ear–nose triangular structure information, neck–shoulder–back–waist–tail linear structure information, and shoulder–waist–forefoot–hindfoot rectangular structure information were extracted in accordance with the skeleton topology. A behavior feature model was established to map different postures. This model consisted of various features, such as coordinate position, direction, speed, angle, area, and length. When a pig was standing, it was judged in accordance with the angle of its head skeleton regardless of whether it raised or lowered its head. The walking behavior was judged in accordance with the rectangular angle of the body. The resting behavior was indicated by the linear expression of the back skeleton when a pig lay down. The evaluation criteria are shown in Figures 9 and 10.



a.Standing b. Walking c. Lying

Figure 9 Overall evaluation criteria



a.Head down b.Head up c.Side face

Figure 10 Local evaluation criteria of pig skeletion

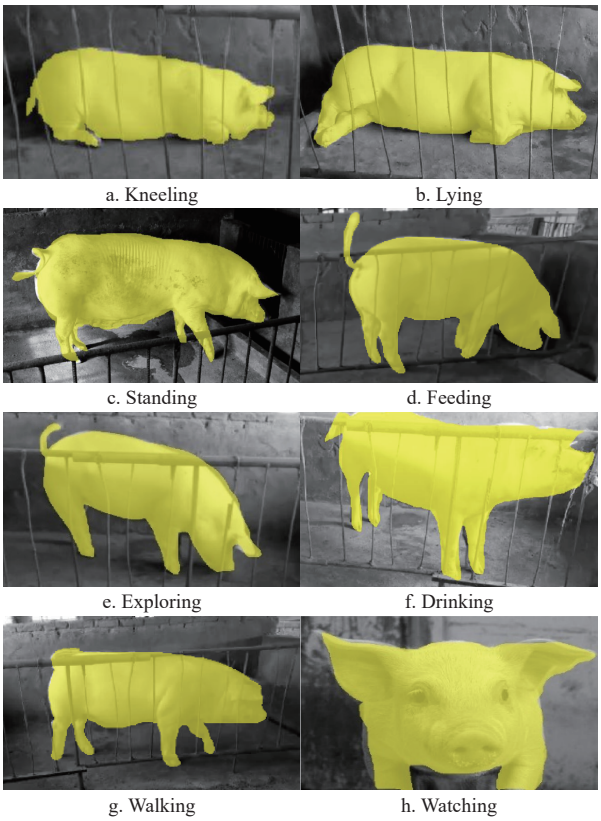
2.4 Body nodes' tracking model

2.4.1 Node Detection Based on DLC

DLC used a neural network in target tracking and is very good at it, but the application scenario is still relatively simple. The DLC algorithm trains images through the ResNet-50 and performs top-down iterative training on image features. The overall target (i.e., a pig) is found, and then the local nodes are found. However, its detection accuracy relies on a large number of manual labeling and massive sample training. Its detection accuracy is low for key points that are not too bright for pigs. We found that there were still tracking drift, loss, and other problems. Especially for the cognition of different parts of the same pig or different pigs, although the training and learning effect had been optimized, the key points such as ears, feet, or snout were often confused. The breakpoint, loss, drift, and other problems existed in the trajectory map. Therefore, we are committed to integrating the minimum frame markup with the maximum tracking efficiency to further improve the robustness and generalization of the algorithm in the pig body nodes' tracking.

First, the training datasets for the pose of individual pigs and groups of pigs were constructed, and the videos of the datasets were shuffled randomly. A dataset primarily consisted of a variety of behavioral videos. Each video was 10 min long. A total of 500 frames of samples were selected. The sequences, coordinates, and affiliations of nodes were used as important reference values for subsequent matching connections. The entire training process used stochastic gradient descent as a network optimizer. We should not only reduce the number of manual marking samples but also enrich the marking information to improve the accuracy of the training model. So we reduced the number of artificial mark images to 10 (the former algorithm required a small sample of 20). Examples of different types of behavior are presented in Figure 11.

Second, the local information about a pig was extremely weak in distinction. The same local area would easily appear in the background and caused confusion. Accordingly, the current study marked the large receptive field area. Occlusion nodes, such as the ear and snout, were evidently more difficult to detect than limb nodes. The head sampling area should be increased to expand the training amount. In view of the problem that the excessive redundancy information of HD video affected the processing accuracy, we compressed the labeled samples 1-2 times unequally and expand the labeled area. In this way, the semantic invariance of the scene is ensured, and the positive correlation information of the training samples is enriched. Moreover, the relatively enlarged sampling points are conducive to reducing the drift displacement of the track points and assimilating the abnormal tracks.



a. Kneeling b. Lying
c. Standing d. Feeding
e. Exploring f. Drinking
g. Walking h. Watching

Figure 11 Different behavior samples of pigs

In addition, data enhancement techniques of spatial coordinate transformation, including rotation, scaling, and hill transformation,

were used to enhance the generalization capability of the network. The coordinate data from different angles obtained via rotation transformation can enhance the cross-perspective recognition capability of the model. Zoom transform changed the size of the coordinates to accommodate different heights and sizes of pigs. Hill transform changed the shape of the coordinates and enhanced the robustness of the model.

Finally, model training was performed. The training was designed to mark seven feature points, namely, the ear, nose, head, shoulder, waist, forefoot, and hindfoot. The spatial probability density of body parts was obtained using the deconvolution layer (DL). The positions of different feature points were extracted through the training network. The tracking model is shown in Figure 12. We generalized the model, with the number of iterations ranging from thousands to millions. At the same time, different batch sizes, shuffles, training rates, and strides were set to optimize the deconvolution layer. The stochastic gradient descent method was adopted as the optimization strategy for model training. The initial learning rate was set to 0.005, and the momentum was set to 0.9. RELU was used as the default activation function for the whole network.

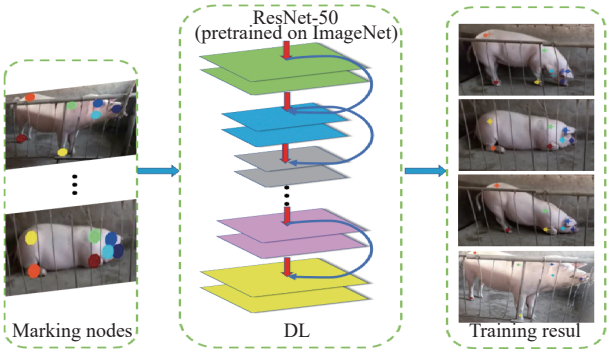


Figure 12 Body nodes' tracking model of DeepLabCut

In this study, based on the traditional DLC, the single high-definition scene in the laboratory was abandoned. In view of the inaccuracy of target tracking, and the loss and drift of trajectory tracking caused by the influence of illumination, ghost shadows, and rapid movement, the original data was preprocessed to expand its label amount and reduce the dependence on artificial labeling. Then, through dimensionality reduction, the scene analysis was guaranteed to remain unchanged, redundant information was eliminated and the trajectory was corrected. Finally, the model was trained to adapt to different scenes, achieving the integration of frame tag dependence minimization and tracking efficiency maximization.

2.4.2 Node Matching Based on PAF

After the detection of nodes, the initial dense skeleton network was built by matching and connecting in accordance with graph theory[31]. This process cannot form an effective description of behavior posture as shown in Figure 13.

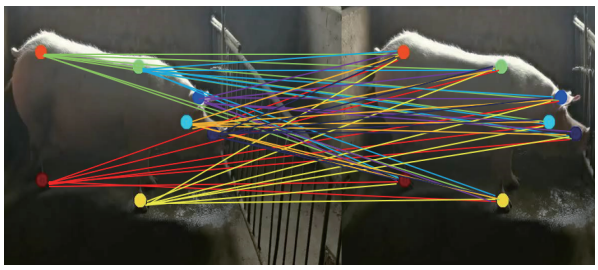


Figure 13 Network of key points

In the current study, two methods were designed to transform the graph structure of the nodes into a sequence. As shown in Figure 14, a network with 14 nodes in a single pig was connected. The first method (i.e., chain sequence) is illustrated in Figure 14a. In accordance with the physical connections, a pig's nodes can be divided into three sequences: head, body, and leg. We concatenated the entire network sequence into a linear sequence. The second method (i.e., traversal sequence) is depicted in Figure 14b. We started from the head node and used the traversal method to visit all the nodes. Then, we output a sequence in accordance with the order of access. Consequently, an initial skeleton was constructed in accordance with the physical order of seven nodes, forming two linear descriptions of head (1-2-3) and body (4-5-6-7).

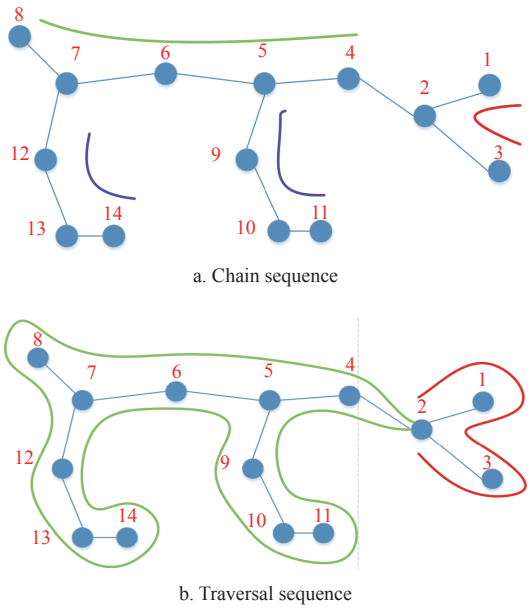


Figure 14 Different node sequences of a single pig

Multi-objective key point connection involves point type and position matching. The position and direction information of PAFs[32] were used to connect multiple pig key points. The connection strategy of nodes among individual pigs is illustrated in Figure 15. Points 1, 2, 3, and 4 were presented as examples, where points 1, 2, and 3 were different key points of the same pig k. The coordinates were X_{j1}, k; X_{j2}, k; and X_{j3}, k. Point 4 was the leg point of another pig $k-1$. V and V_t represented unit and vertical vectors, respectively.

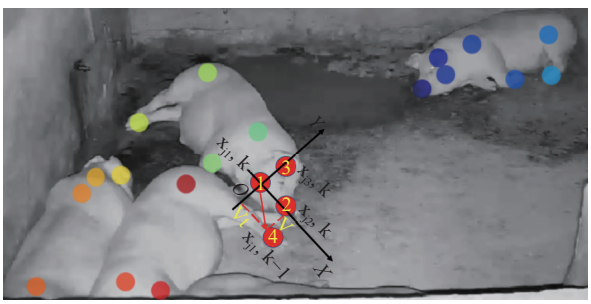


Figure 15 Connection strategy of nodes among individual pigs

With regard to $L_k^*(P)$ as the PAF of a pig:

$$L_k^*(P) = \begin{cases} V\left(\text{point } P \text{ belongs to pig } k\right) \\ O\left(\text{point } P \text{ does not belong to pig } k\right) \end{cases} \quad (15)$$

where, O is the zero vector. For any point P in the region, the value

of $L_k^*(P)$ is the unit vector V at that point. $L_k^*(P)$ is the zero vector when P is in any other position.

However, the actual pig region was composed of numerous points, and any point P was required to be less than the distance threshold. The following conditions must be met:

$$\begin{cases} 0 \le \left| V\left(P - X_{j1,k}\right) \right| \le l_k \\ \left| Vt\left(P - X_{j1,k}\right) \right| \le \sigma_l \end{cases} \tag{16}$$

where, l_k is the length of pig parts between adjacent nodes; σ_l is the width of pig parts between adjacent nodes.

If multiple targets k overlapped at point P, then the vector field was the mean of all the target vector fields. Point $P(u)$ was taken on the directrix between adjacent nodes after obtaining the vector at each point P in the pig region.

$$P(u) = (1-u)d_{j1} + ud_{j2} \tag{17}$$

where, $P(u)$ is the sampling point on the node. d_{j1} is the predicted coordinate of node j1. d_{j2} is the predicted coordinate of node $j2$. u is the relative distance between d_{j1} and d_{j2}, and its value is 0–1. The partial affinity vector $L(P(U))$ predicted at the position of each point $P(u)$ was multiplied by the unit vector V from j_1 to j_2 and then summed.

The body weight calculation formula for the connecting nodes j_1 and j_2 was as follows:

$$E = \int_0^1 L(P(u)) \frac{d_{j2} - d_{j1}}{\left\| d_{j2} - d_{j1} \right\|^2} d_u. \tag{18}$$

When the partial affinity vector at position $P(u)$ was in the same direction as unit vector V, the weight E was the maximum. Given that countless points occurred on the line segment between adjacent nodes, equal interval sampling was performed on the line segment during practical operation to calculate the weight E of $P(u)$ points.

The predicted candidate point set D_J was as follows:

$$D_j = \{D_{jm} : j \in \{1,2,3,\cdots,J\}, m \in \{1,2,3,\cdots,J\}\} \tag{19}$$

where, j is the type number of a node, m is the sequence number of a node; D_{jm} is the mTH node of type j.

The node sets D_{j1} and D_{j2} were regarded as graph nodes, and the connections of node sets were regarded as graph edges. Then, the weight E of PAF was evaluated to find the edge with the highest weight. The sum of the body weights of class c was represented by E_c, and the formula was as follows:

$$\max_{Z_c} E_c = \max_{Z_c} \sum_{m \in D_{j1}} \sum_{n \in D_{j2}} E_{mn} z_{j1}^m z_{j2}^n \tag{20}$$

where, z_{j1}^m and z_{j2}^n are the connection states between two points, which are 0 or 1. Z_c is a subset of class c. Emn is the weight of PAF between nodes D_{j1} and D_{j2}. D_{j1} is the node set of class $j1$. D_{j2} is the node set of class $j2$.

Two edges connected to the same node must be as follows:

$$\begin{cases} \sum_{n \in D_{j2}} z_{j1}^m z_{j2}^n \le 1 \left(m \in D_{j1}\right) \\ \sum_{m \in D_{j1}} z_{j1}^m z_{j2}^n \le 1 \left(n \in D_{j2}\right) \end{cases} \tag{21}$$

Therefore, whether the maximum weights of nodes 3 and 4 conform to class c can be determined on the basis of nodes 1 and 2.

3 Results and discussion

3.1 Results of skeleton extraction

Pigs walk with a nonrigid motion, and thus, their limbs or head will be deformed, resulting in missing skeleton endpoints. First, the distribution of pixel points in an image was detected to establish the eight neighborhoods of target pixel points. Non-skeleton points were deleted to realize single-pixel processing by determining whether deleting the center pixel will affect the connectivity of the skeleton. Consequently, the burrs near the intersection and the endpoints will be reduced. The skeleton extraction results of eight types of behavior are presented in Figure 16.

3.2 Behavior posture database

3.2.1 Extraction of body nodes

To construct the standard posture database, a node model with common characteristics must be established. After extracting the main nodes with different types of behavior, redundant nodes remained, affecting the skeleton structure and resulting in description errors. Therefore, different nodes were reserved for different types of behavior, and secondary nodes were eliminated. The optimization of nodes with different types of behavior is illustrated in Figure 17. The limb nodes of the kneeling posture were removed. The forefoot of the lying posture was removed. The tail node was removed from the standing and eating postures. The tail and neck nodes were removed from the walking posture. The drinking posture preserved seven main nodes. The watching posture retained the center point. The optimized skeleton exhibited more feature differentiation, which was beneficial for enriching the posture database samples.

3.2.2 Posture database of pigs

The initial posture database only consisted of typical samples of different types of behavior. The skeleton will have many variants due to the nonrigid body movement of pigs. We supplemented the samples as much as possible through spatial data enhancement technology to improve the richness of skeleton mapping. Skeleton behavior samples are shown in Figure 18.

3.3 Results of body node tracking

3.3.1 Tracking results of an individual pig

In the current study, seven body nodes of individual sows were extracted. The results showed that the algorithm can continuously and effectively extract different behavioral nodes of pigs. For example, Figure 19 shows the tracking results of two video sequences from different angles, basically realizing all-weather detection.

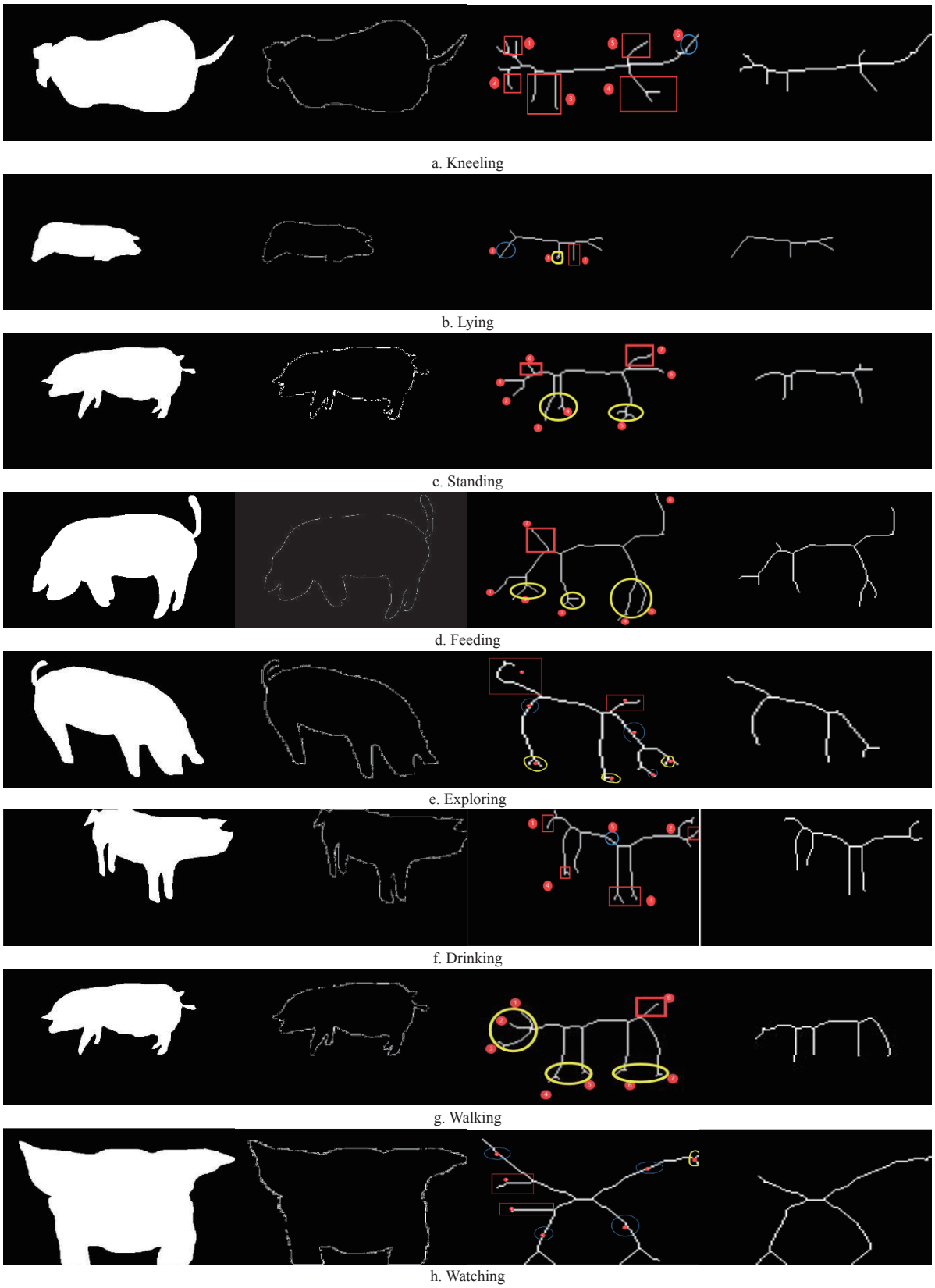
3.3.2 Tracking Results of Piglets

The current study also conducted exploratory experiments on multi-objective scenes during the day and at night. It determined that the main behavior of the four piglets, such as standing and lying, can be tracked successfully under the condition of less movement of the piglets at night. However, the piglets produced considerable occlusion and interaction during the day. The algorithm worked well in the case of clear individual intervals. Problems, such as tracking loss and drift, occurred when pigs gathered and moved quickly. The algorithm should be further optimized under such conditions to improve its robustness. The tracking results of two different scenes are shown in Figure 20.

3.4 Discussion

3.4.1 Comparison of skeleton extraction algorithms

In order to verify the generalization ability of the algorithm, the public data set COCO2014 was supplemented for experimental comparison. In addition, there are images of numbers and people. The results are presented in Figure 21. The proposed algorithm was compared with MR, DT, MAT, and the traditional ZS method. A large amount of redundant information and burr could be found in MR. DT optimized redundant information and found the central



a. Kneeling

b. Lying

c. Standing

d. Feeding

e. Exploring

f. Drinking

g. Walking

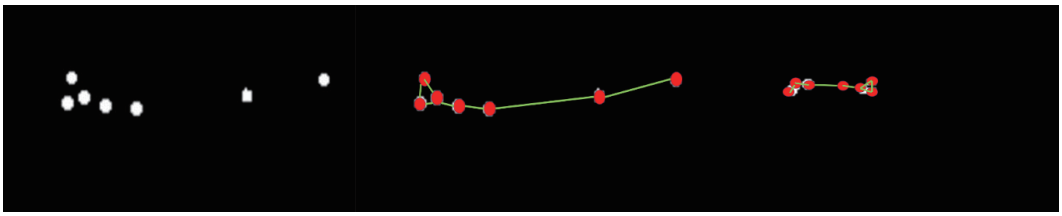
h. Watching

Figure 16 Results of skeleton extraction and burr correction

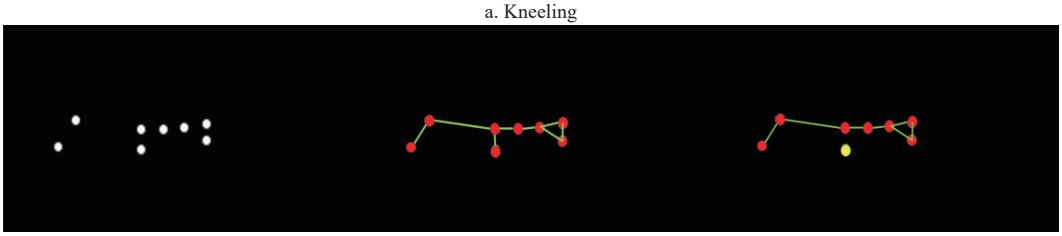
axis, but it still required optimization. MAT had break points. The traditional ZS algorithm considerably improved the processing capability of redundant information and burr. On the basis of this algorithm, the proposed algorithm further optimized the removal capability of boundary burrs and redundant branches. It was not only robust in the aspects of pig skeleton extraction and burr removal, but it is also the best for digital and human skeleton thinning among various methods.

3.4.2 Performance of Body Node Matching

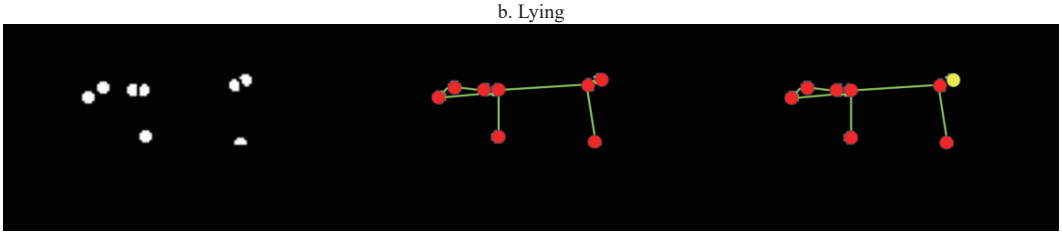
The connection settings of the seven feature points in the current study are listed in Table 3. The matching performance experiment of body nodes based on PAF is illustrated in Figure 22. Figure 22a shows the connection effect of individual pig nodes that basically realized the accurate estimation of posture, and only



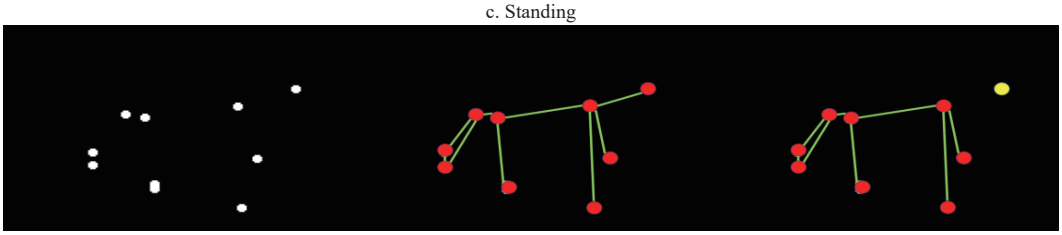
a. Kneeling



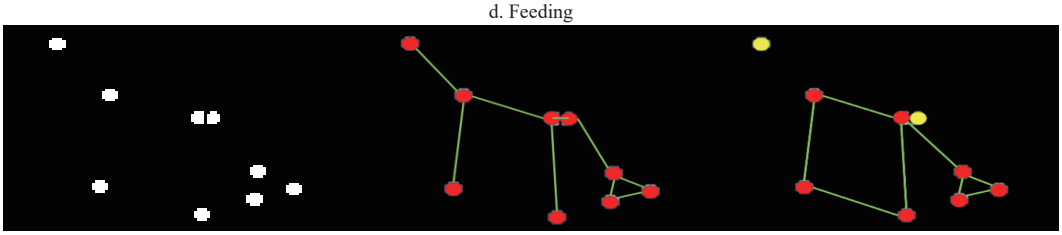
b. Lying



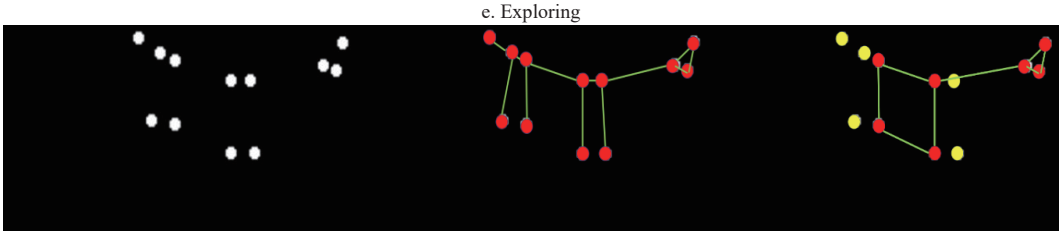
c. Standing



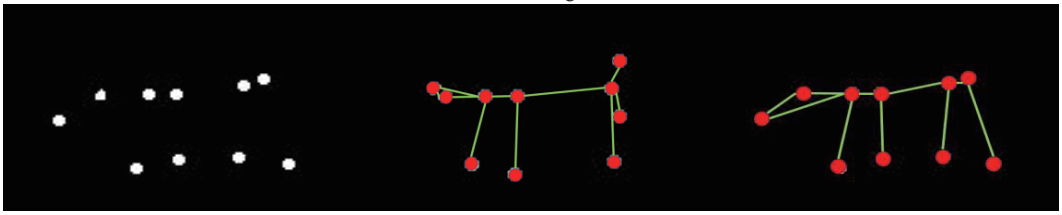
d. Feeding



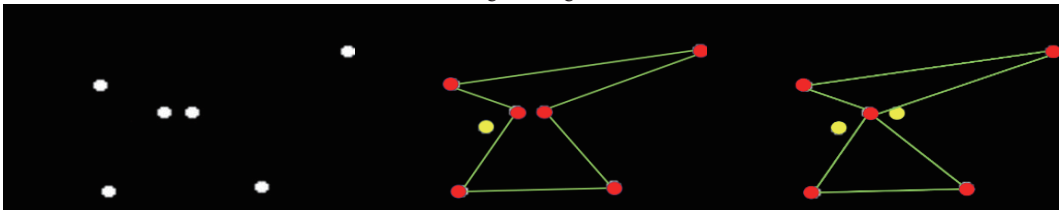
e. Exploring



f. Drinking



g. Walking



h. Watching

Figure 17 Extraction and optimization of node points

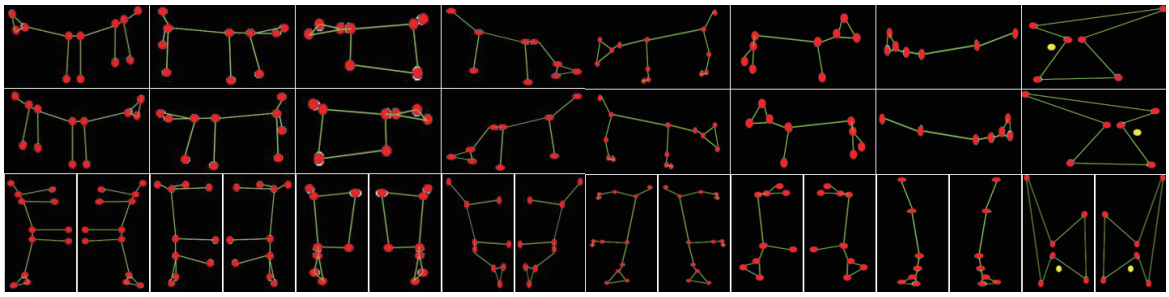


Figure 18 Behavior posture database of pigs

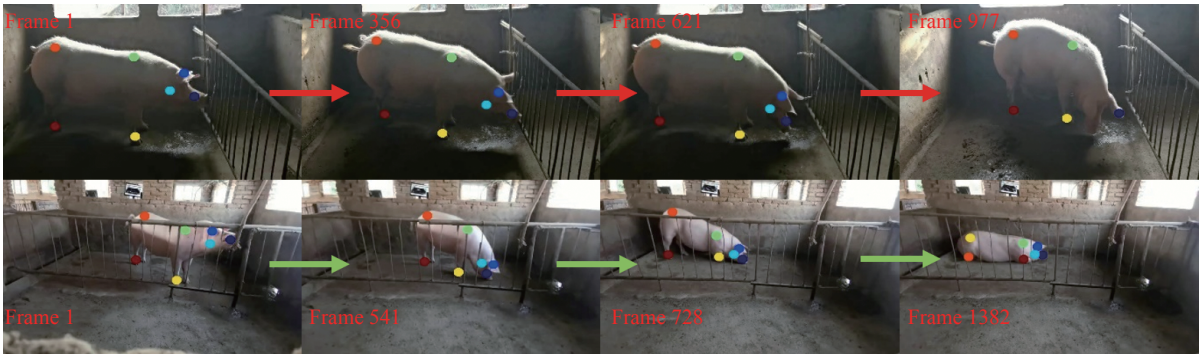


Figure 19 Tracking results of an individual pig

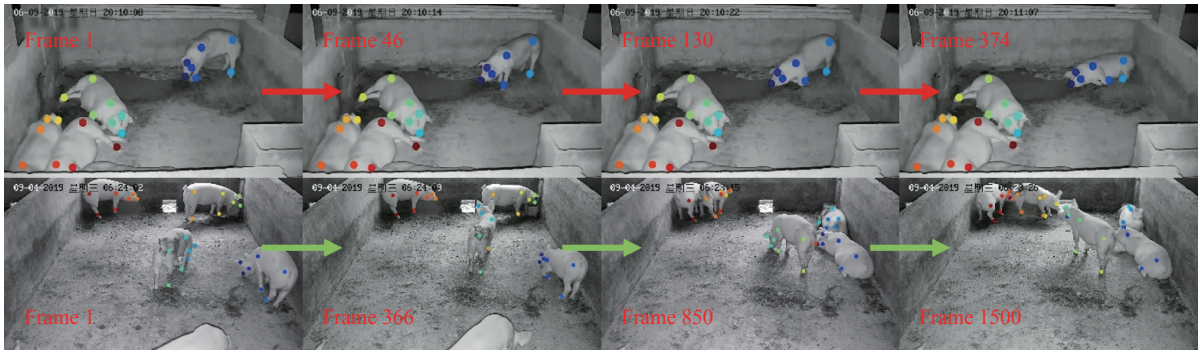
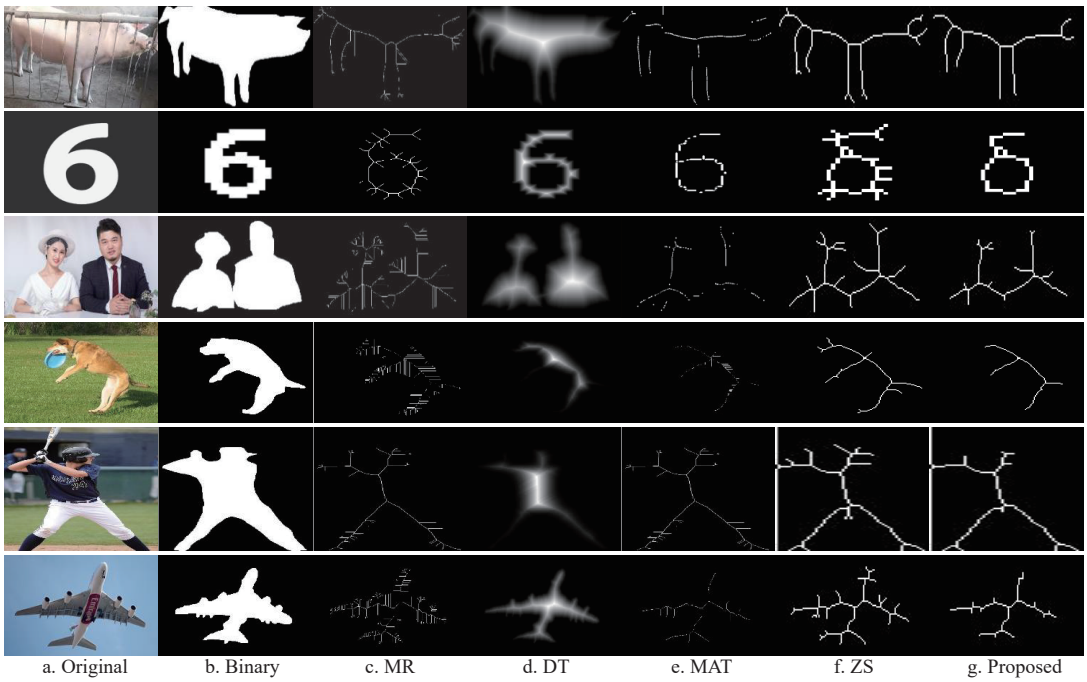


Figure 20 Tracking results of piglets



a. Original b. Binary c. MR d. DT e. MAT f. ZS g. Proposed

Note: MR: morphology refinement; DT: Distance Transformation; MAT: Medial Axis Transformation; ZS: Zhang-Suen thinning algorithm.

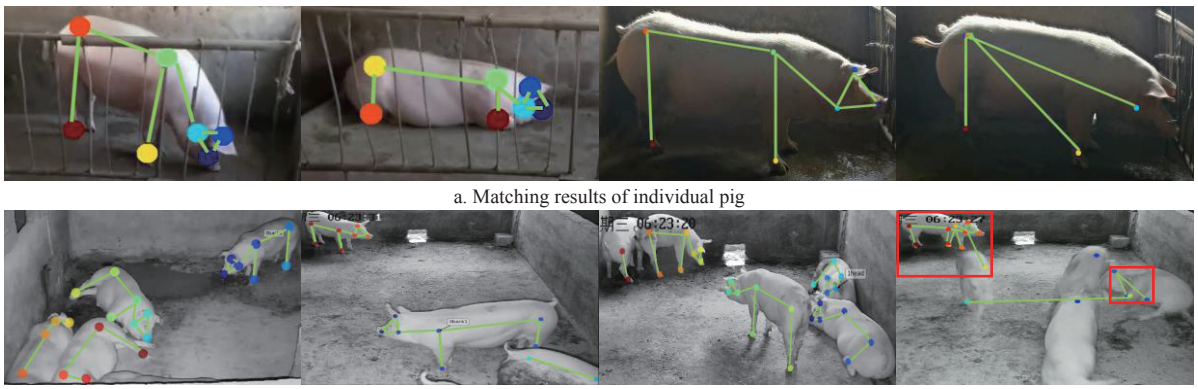
Figure 21 Comparison of skeleton extraction algorithms

Table 3 Custom feature points and connection lines

No.	Name	Connection lines
1	ear	(1-2), (1-3)
2	nose	(2-1), (2-3)
3	head	(3-1), (3-2), (3-4)
4	shoulder	(4-3), (4-5), (4-6)
5	waist	(5-4), (5-7)
6	forefoot	(6-4)
7	hindfoot	(7-5)

temporary drift loss occurred in the tracking of head nodes. Figure 22) shows the connection effect of piglets. The effect was decreased compared with that of individual pigs due to the difficulty in multi-target tracking and occlusion problems. The probability of tracking loss and matching failure was increased. But for general occlusion and other problems have been significantly improved. Further improvement can be made in the robustness of the matching model.

The accuracy of node matching depended on accurate target tracking. The overall tracking accuracy was reduced to a certain extent due to the addition of group targets to be tracked in the current study. Training error can be reduced to 1.4 pixels, and the test error of high-definition images was 13.74 pixels. The tracking accuracy of the algorithm in the current study reached 85.08%. The matching effect of the nodes was still affected due to piggery density and individual interaction, and thus, should be further improved. The algorithm loss value can be reduced to 0.007, and the learning rate was 0.002, as shown in Figure 23.



a. Matching results of individual pig

b. Matching results of piglets

Figure 22 Matching performance of body nodes of pigs

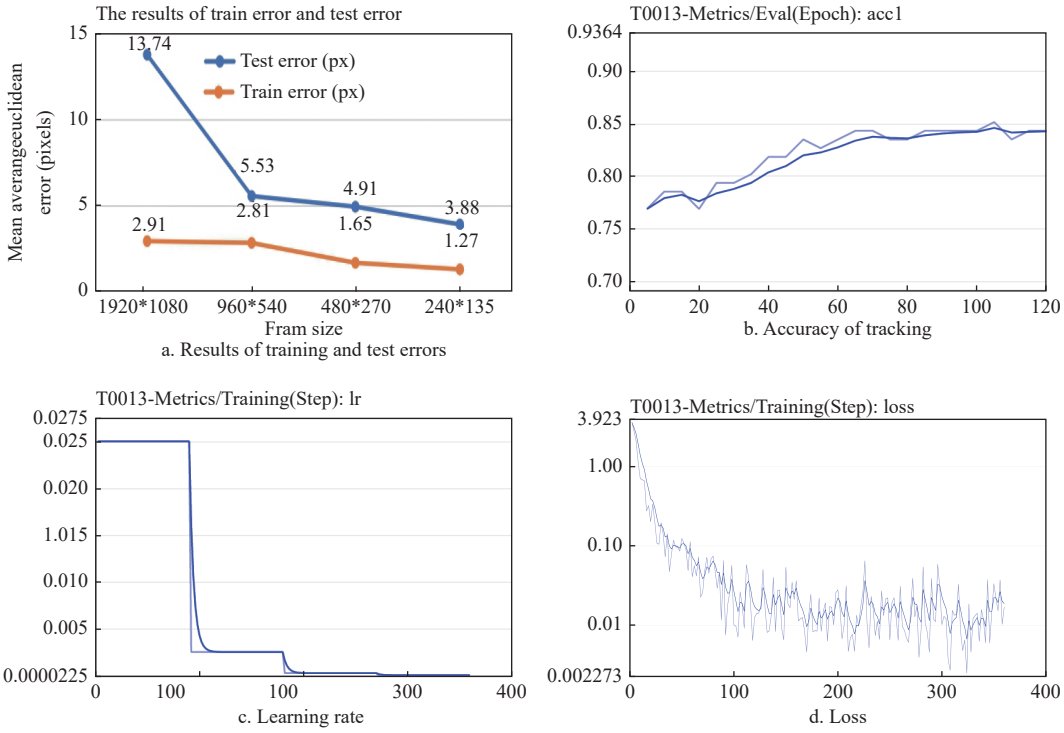


Figure 23 Algorithm performance of the proposed method of this study

3.4.3 Behavior Posture Classification

This study matched the behavior posture with the standard posture database by using SVM. The matching results obtained in Section 3.4.2 were used as the input data of SVM and compared with the skeleton pose database obtained in Section 3.2.2 to finally realize the classification of different behaviors of pigs. It realized different behavior classifications by integrating the posture evaluation standard, such as skeleton head shape, length, height, angle, and coordinates. Meanwhile, Table 4 lists the classification prediction results. The sample number of each behavior was 2000.

The confusion matrix is shown in Figure 24. Although the specific subclassification results were approximately 70%, the primary reason was that the main types of pig posture were standing and resting. The proportion of broad classification can reach approximately 96%, indicating that the proposed algorithm and the extracted skeleton pose were distinguishable.

Table 4 Predictions of different types of behavior

Predictions	Kneeling	Lying	Standing	Feeding	Exploring	Drinking	Walking	Total
Kneeling	1406	558	13	4	16	0	3	2000
Lying	481	1265	43	7	72	45	87	2000
Standing	0	47	1279	191	65	112	306	2000
Feeding	0	3	147	1358	266	116	110	2000
Exploring	2	2	46	355	1210	7	378	2000
Drinking	2	0	192	34	35	1533	204	2000
Walking	9	35	240	101	186	387	1042	2000

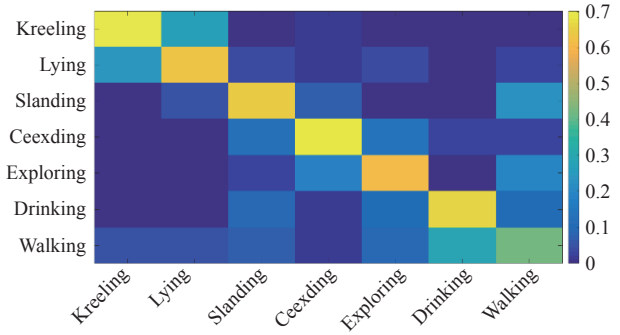


Figure 24 Confusion matrix of the behavior posture classification of pigs

The average accuracy (ACC) of behavior classification based on SVM in the current study was 89.6%. Other evaluation indexes are listed in Table 5. P represents precision rate, R represents recall rate, F1 represents the F1-score, F1-score is the harmonic mean of precision rate and recall rate; ACC represents accuracy rate, TPR represents true positive rate, and FPR represented false positive rate.

Table 5 Evaluation indicators of different behavior categories

Behavior	P	R	F1	ACC	TPR	FPR
Kneeling	0.74	0.703	0.721 026	0.922 286	0.703	0.041 167
Lying	0.662 304	0.6325	0.647 059	0.901 429	0.6325	0.053 75
Standing	0.652 551	0.6395	0.645 96	0.899 857	0.6395	0.056 75
Feeding	0.662 439	0.679	0.670 617	0.905 361	0.682 412	0.057 667
Exploring	0.654 054	0.605	0.628 571	0.897 857	0.605	0.053 333
Drinking	0.696 502	0.7665	0.729 826	0.905 986	0.696 818	0.055 667
Walking	0.489 202	0.521	0.5046	0.841 831	0.473 636	0.090 667
Average	0.651 007	0.6495	0.649 666	0.896 372	0.633 267	0.058 429

The results showed that walking behavior was more difficult to distinguish compared with other types of behavior. Meanwhile, kneeling behavior achieved the highest recognition rate due to its simple and unique skeleton. In addition, the different subcategories of behavior interfered with one another considerably, and the model evaluation parameters and posture database should be increased to improve the precision of classification. The small sample exploration experiment in the current study proved that the research was feasible and played a supporting role in the analysis of different types of animal behavior and posture.

4 Conclusions

On the basis of the multi-pig skeleton extraction and pose estimation method of ZS-DLC-PAF, the current study extracted the skeleton information of pigs in real time, constructed a pig behavior posture database, and conducted effective pose estimation.

First, to solve the problem of redundant information and burr in pig skeleton extraction, redundant information deletion templates were designed and a burr chain code mechanism was introduced to achieve skeleton pixel simplification. Then, on the basis of DLC and PAF, accurate tracking and connection of multi-pig skeleton nodes were realized. Finally, in accordance with the posture evaluation criteria, SVM was used to match the pig posture database to realize the different classifications of pig behavior. The posture database can provide technical support to pig behavior recognition. The method used in the current study improved the accuracy of body node tracking and connection. The breakthrough and innovation in the generalization and robustness of this method were verified from the perspectives of quantitative analysis and qualitative evaluation.

Given the high complexity and different backgrounds of the acquired target images, the future skeleton extraction algorithm can be optimized from two aspects, i.e., by improving the tracking accuracy of multi-target joints and by expanding the posture database.

Acknowledgments

This work was financially supported by the National Major Science and Technology Project (Innovation 2030) of China (Grant No. 2021ZD0113701).

[References]

[1] Nasirahmadi A, Edwards S A, Matheson S M, Sturm B. Using automated image analysis in pig behavioural research: Assessment of the influence of enrichment substrate provision on lying behaviour. Applied Animal Behaviour Science, 2017, 196: 30–35.

[2] Naseri M, Heidari S, Gheibi R, Gong L H, Sadri A. A novel quantum binary images thinning algorithm: a quantum version of the Hilditch's algorithm. Optik-International Journal for Light and Electron Optics, 2016; 131: 678–686.

[3] Chen C, Zhu W X, Norton T. Behaviour recognition of pigs and cattle: journey from computer vision to deep learning. Computers and Electronics in Agriculture, 2021; 187: 106255.

[4] Kustra J, Jalba A, Telea A. Computing refined skeletal features from medial point clouds. Pattern Recognition Letters, 2016; 76: 13–21.

[5] Gronskyte R, Clemmensen L H, Hviid M S, Kulahci M. Pig herd monitoring and undesirable tripping and stepping prevention. Computers and Electronics in Agriculture, 2015; 119: 51–60.

[6] Nasirahmadi A, Hensel O, Edwards S A, Sturm B. Automatic detection of mounting behaviours among pigs using image analysis. Comput. Electronics in Agriculture, 2016; 124: 295–302.

[7] Nasirahmadi A, Edwards S A, Sturm, B. Implementation of machine vision for detecting behaviour of cattle and pigs. Livestock Science, 2017; 202: 25–38.

[8] Kusuma W A, Husniah L. Skeletonization using thinning method for human motion system. 2015 International Seminar on Intelligent Technology and Its Applications (ISITIA), Surabaya: IEEE, 2015; pp.103-106. doi: 10.1109/ISITIA.2015.7219962

[9] Zhang T Y, Suen C Y. A fast parallel algorithm for thinning digital patterns. Communications of the ACM, 1984; 27(3): 236–239.

[10] Ramya P, Rajeswari R. Human action recognition using distance transform and entropy based features. Multimedia Tools and Applications, 2021; 80(21): 8147–8173.

[11] Shi C W, Zhao J Y, Chang J S. Skeleton feature extraction algorithm based on medial axis transformation. Computer Engineering, 2019; 45(7): 242–250. (in Chinese)

[12] Lynda B B, Basel S, Abdelkamel T. A modified ZS thinning algorithm by a hybrid approach. The Visual Computer, 2018; 34(5): 689–706.

[13] Lynda B B, Basel S, Abdelkamel T. Implementation and comparison of binary thinning algorithms on GPU. Computing, 2018; 101(8): 1091–1117.

[14] Li R, Zhang X Y. Research on the improvement of EPTA parallel thinning algorithm. Proceedings of the 2018 International Conference on Network, Communication, Computer Engineering (NCCE2018), 2018; pp.994–1001. doi: 10.2991/ncce-18.2018.167

[15] Pfister, T, Charles J, Zisserman A. Flowing convnets for human pose estimation in videos. 2015 IEEE International Conference on Computer Vision (ICCV), Santiago: IEEE, 2015; pp.1913–1921. doi: 10.1109/ICCV.2015.222

[16] Newell A, Huang Z A, Deng J. Associative embedding: end-to-end learning for joint detection and grouping. In: Proceedings of the 31st International Conference on Neural Information Processing Systems, 2017; pp.2274–2284. doi: 10.5555/3294771.3294988.

[17] Fang H S, Xie S Q, Tai Y W, Lu C W. RMPE: Regional multi-person pose estimation. In: 2017 IEEE International Conference on Computer Vision (ICCV), Venice: IEEE, 2017; pp.2353–2362. doi: 10.1109/ICCV.2017.256.

[18] Liao R J, Cao C S, Garcia E B, Yu S Q, Huang Y Z. Pose-based temporal-spatial network (PSTN) for gait recognition with carrying and clothing variations. In: Proceedings of the 12th Chinese Conference on Biometric Recognition (CCVR 2017), 2017; pp.474–483. doi: 10.1007/978-3-319-69923-3_51.

[19] Cowton, J., Kyriazakis, I., Bacardit, J. Automated individual pig localisation, tracking and behaviour metric extraction using deep learning. IEEE Access, 2019; 7: 108049–108060.

[20] Gan H M, Ou M Q, Zhao F Y, Xu C G, Li S M, Chen C X, et al. Automated piglet tracking using a single convolutional neural network. Biosystems Engineering, 2021; 205(1): 48–63.

[21] Gan H M, Ou M Q, Huang E D, Xu C G, Li S Q, Li J P, et al. Automated detection and analysis of social behaviors among preweaning piglets using key point-based spatial and temporal features. Computers and Electronics in Agriculture, 2021; 188: 106357.

[22] Gan H M, Li S M, Ou M Q, Yang X F, Huang B, Liu K, et al. Fast and accurate detection of lactating sow nursing behavior with CNN-based optical flow and features. Computers and Electronics in Agriculture, 2021; 189: 106384.

[23] Mathis A, Mamidanna P, Cury K M, Abe T, Murthy V N, Mathis M W, et al. DeepLabCut: Markerless pose estimation of user-defined body parts with deep learning. Nature Neuroscience, 2018; 21(9): 1281–1289.

[24] Nath T, Mathis A, Chen A C, Patel A, Bethge M, Mathis M W. Using DeepLabCut for 3D markerless pose estimation across species and behaviors. Nature Protocols, 2019; 14(7): 476531.

[25] Alameer A, Kyriazakis I, Bacardit J. Automated recognition of postures and drinking behaviour for the detection of compromised health in pigs. Scientific Reports, 2020; 10: 13665.

[26] Cheng F, Zhang T M, Zheng H K, Huang J D, Cuan K X. Pose estimation and behavior classifcation of broiler chickens based on deep neural networks. Computers and Electronics in Agriculture, 2021; 180: 105863.

[27] Romero-Ferrero F, Bergomi M G, Hinz R C, Heras F J H, de Polavieja G G. Idtracker. ai: Tracking all individuals in large collectives of unmarked animals. Nature Methods, 2019; 16: 179–182.

[28] Sun S J, Akhtar N, Song H S, Mian A, Shah M. Deep affinity network for multiple object tracking. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019; 43(1): 104–119.

[29] Zhang Y F, Wang C Y, Wang X J, Zeng W J, Liu W Y. A simple baseline for multi-object tracking. International Journal of Computer Vision, 2021; 129(11): 3069–3087.

[30] Jiang Y Q, Wang P, Gao H W, Jin L, Liu X J. Study on the method for removing boundary burr based on relevance of chain code. In: 2011International Conference on Electronice Commerce, Web Application and Communication (ECWAC 2011), 2011; 144: 188–194.

[31] Kipf T N, Welling M. Semi-supervised classification with graph convolutional networks. N: The 5th International Conference on Learning Representations, 2016. arXiv: 1609.02907.

[32] Cao Z, Simon T, Wei S E, Sheikh Y. Realtime multi-person 2D pose estimation using part affinity fields. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu: IEEE, 2017; pp.1302–1310. doi: 10.1109/CVPR.2017.143.