Computers and Electronics in Agriculture 206 (2023) 107666

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

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

Automatic estimation of dairy cow body condition score based on attention-guided 3D point cloud feature extraction

Wei Shi [a], Baisheng Dai [a,*], Weizheng Shen [a,*], Yukun Sun [b], Kaixuan Zhao [c], Yonggen Zhang [b]

[a] College of Electrical Engineering and Information, Northeast Agricultural University, Harbin 150030, China
[b] College of Animal Science and Technology, Northeast Agricultural University, Harbin 150030, China
[c] College of Agricultural Equipment Engineering, Henan University of Science and Technology, Luoyang 471023, China

ARTICLE INFO

Keywords:
Dairy cow
Body condition score
Point cloud feature extraction
Attention mechanism
Deep learning

ABSTRACT

Body condition score (BCS) is an important indication in management of dairy cow breeding, which can be used to evaluate milk production, reproduction and health status of dairy cows. Due to the low efficiency and subjectiveness of manual scoring, it's necessary to estimate BCS of dairy cows automatically and accurately. The current automatic estimation methods generally utilized features extracted from 2D image data or manually defined 3D surface features of dairy cows' back end, which were defective or difficult to essentially represent the 3D concavity information of local area of dairy cattle body severed as the key indicator of BCS. In this study, a 3D data format dataset was first built for dairy cow body condition score estimation. To learn a more effective representation of concavity information automatically and focus on vision saliency information precisely of local area of dairy cows' back end, an automatic method was proposed to estimate BCS of dairy cows based on attention-guided 3D point cloud feature extraction. Experiments show that proposed body condition scoring model achieved the accuracy of 0.49, 0.80, 0.96 within 0, 0.25, 0.50 point deviation respectively, which has achieved good estimation results in comparison with the other research.

1. Introduction

Body condition score is an important tool for management of dairy cow, which can reflect the relative amount of subcutaneous fat as energy reserve of dairy cows (Azzaro et al., 2011). Its assessment is usually based on appearance of tissue coverage on bovine back and pelvic bone protrusions (Ferguson et al., 1994; Wildman et al., 1982; Spoliansky et al., 2016; Liu et al., 2020). Body condition score is generally used 5-point scale system (Ferguson et al., 1994; Spoliansky et al., 2016), with an increment of 0.25. Particularly, 1 point of BCS represents dairy cow is extremely thin, 5 point of BCS represents dairy cow is severely obese. In the management process of modern farms, BCS can judge whether nutrient distribution of dairy cows is reasonable, which can also be used to evaluate milk yield, reproduction and health status (Anglart, 2010), etc. Therefore, the estimation of dairy cow body condition is of great significance to the development of precision animal husbandry.

Manual scoring method by trained scorer is strongly subjective, poorly repeatable, and less efficient, which make it difficult to adapt to the development of modern animal husbandry (Bercovich et al., 2013; Alvarez et al., 2018). With the development of intelligent sensing technology and computer vision technology, non-contact automatic scoring methods of dairy cow body condition have been researched. According to the different ways of feature extraction, non-contact methods can be divided into manual feature extraction based and deep learning feature extraction based scoring methods. While manual feature extraction based methods can further be divided into 2D feature extraction based and 3D feature extraction based. Methods based on 2D feature extraction generally used visual sensors such as RGB and thermal camera to capture images of each dairy cow's region of interest. Then image processing, data statistics and analysis were used to measure the degree of convexity, body length and other indicators of key parts of dairy cow to perform regression analysis to obtain BCS. Bewley et al. (2008) captured top-view RGB images of dairy cows and manually marked anatomical points on images to describe the outline of dairy cow, based on which body condition related features were extracted, such as hook angle, tail concavity and etc. Then a regression model was constructed to estimate BCS. Halachmi et al. (2013) localized the contour of dairy cows in thermal images and extracted features based on deviation of the contour from its fitted parabola to estimate BCS. While methods based on 3D feature exaction mainly combined 3D sensor with

* Corresponding authors.
E-mail addresses: bsdai@neau.edu.cn (B. Dai), wzshen@neau.edu.cn (W. Shen).

https://doi.org/10.1016/j.compag.2023.107666
Received 20 March 2022; Received in revised form 5 August 2022; Accepted 17 January 2023
Available online 27 January 2023
0168-1699/

data processing technology to extract convex hull volume, mean height, roughness of 3D surface and the other 3D feature to build a BCS regression model. Zin et al. (2020) selected two regions of interest from 3D surface data of the back of dairy cow. With respect to the first area, body condition related features such as average height, volume of the convex hull, roughness of 3D surface and difference between convex hull and volume are extracted, and for the second area, features such as average height and extreme value are specially extracted. Based on these features extracted from two area, different regression models were designed and compared for BCS estimation. Liu et al. (2020) defined 3D shape features from the selected six regions within the defined ROI formed with the landmarks on the cow's back end and used an ensemble learning approach to estimate BCS of dairy cow. Although above-mentioned body condition scoring methods based on manual feature extraction have achieve good results, those features are manually defined and selected, which are empirical and subjective and thus not robust for BCS estimation in complex scenarios.

To extract features representing body condition more effectively and automatically, deep learning based methods usually built an end to end BCS estimation model by learning body condition related features with less or without manual intervention. Alvarez et al. (2018) proposed a first deep learning based BCS estimation model, which used the light-weight SqueezeNet network to learn features directly from a specific image data consisted of the depth image of dairy cow and its edge detection and Fourier transform results. Sun et al. (2019) employed a DenseNet-BC network to learn body condition related features for an end to end BCS estimation, based on the image consisted of depth, gray and phase consistency of dairy cow image. These methods improved the performance of BCS estimation benefiting from the effective feature representation of deep learning model. However, the feature extracted by these methods mainly learned from image data, although consisting of the depth image, which was difficult to essentially represent the 3D concavity information of dairy cows' back end. In light of manual estimation of BCS by trained scorer with naked eyes judging from 3D shape of dairy cow, it is necessary to explore the potential of learning body condition related features from 3D data directly and automatically.

In this study, point cloud, a kind of 3D data, was specially considered to learn 3D shape features for an automatic estimation of BCS. Comparing to other 3D data, such as mesh, voxel, and etc., point cloud is the simplest 3D representation method, which is close to raw sensor data and can represent 3D shape more sparsely and accurately (Qi et al., 2017b). To learn effective 3D features from point cloud, a feature extraction framework based on point cloud was first introduced in this work. More importantly, to facilitate and guide the framework extracting more valuable features from those local areas which reflecting concavity information more significantly, a vision attention mechanism was particularly designed and integrated into the feature extraction framework. The main contributions of this work are: Firstly, a series of data processing strategies of point cloud of dairy cow was presented to remove redundant information and normalize point cloud of different dairy cows. Secondly, a point cloud dataset of 512 dairy cows' back end was built. Thirdly, an attention-guided 3D point cloud feature extraction model for the purpose of automatic BCS estimation was proposed. Fourthly, a competitive performance of BCS estimating was achieved.

The rest of this work are as follows: Section 2 describes the collection of dataset and data processing of point cloud, and the details of feature extraction network and designed attention guidance mechanism. Section 3 provides a comparative analysis and discussion of different experimental results. Section 4 concludes this study and provides an outlook for future work.

2. Materials and method

The main workflow of this work is summarized in Fig. 1. A 3D data acquisition unit was designed and used to collect experimental data. Data processing strategies, including noise point cloud removal, ROI extraction and translation, down-sampling and normal vector addition, was present to acquire more instructive point cloud data. An attention-guided 3D point cloud feature extraction was then proposed to learn body condition related features and estimate BCS. Details of this workflow are described as follows.

2.1. Data collection

The collection of experimental data was carried out on two standardized farms, Nestlé Farm and Yuanzhongyuan Farm, which

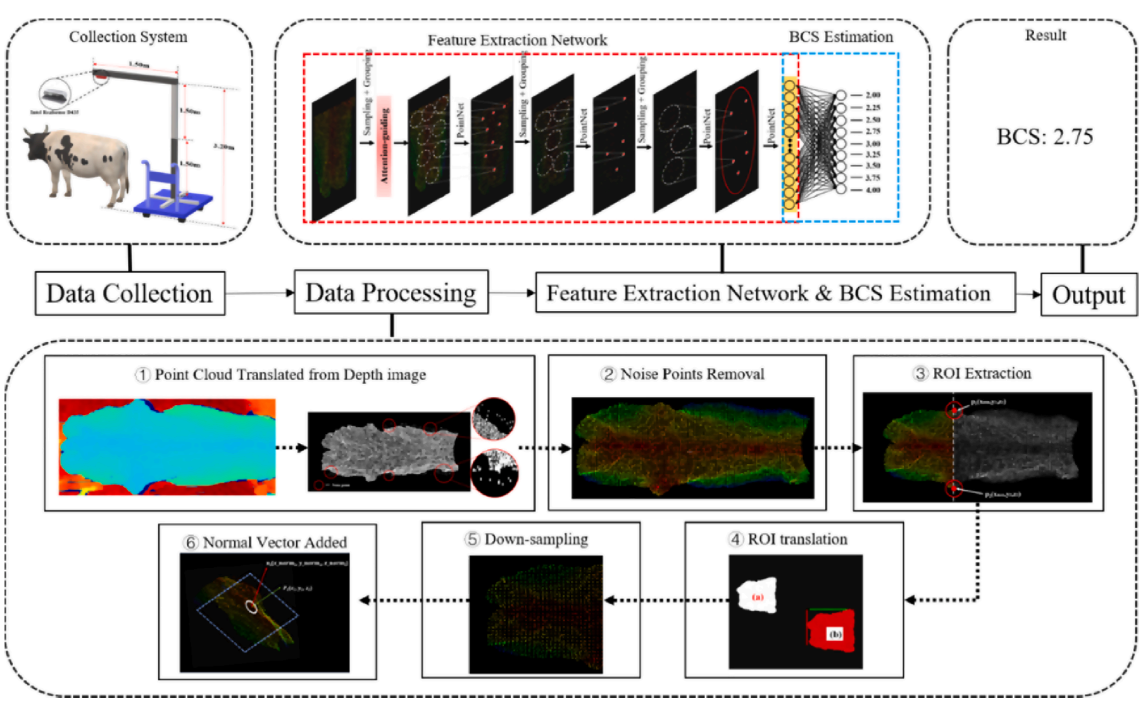


Fig. 1. The workflow of this study.

respectively located in Harbin City and Anda City, Heilongjiang Province. Due to the grayscale value of each pixel point of depth image can be used to characterize the distance of a point in the scene from the camera, and the geometry of the visible surface of the scene is directly reflected. Therefore, depth image can be computed as point cloud data after a coordinate transformation. In this experiment, a total of 512 dairy cows were randomly selected from two farms to collect the depth image data of dairy cow, and a total of 3,660 images were collected. Then, this study used depth information contained in acquired depth image, that is, the depth frame data, combined with the built-in parameters of the camera to generate point cloud of dairy cow, and use the result as original point cloud data and save it in “. *pcd*” format,

In order to obtain top-view depth image of dairy cow, the visual sensor used in data acquisition device in this experiment is Intel Realsense D435 camera, which has three cameras of color, infrared and depth. And the resolution of depth camera is 1280 × 720 pixels, the applicable height range of the camera is 1.5 m ~ 5 m. In addition, the camera also has a smaller volume, which cause it convenient to install, and don’t need a separate power supply. In this experiment, three 30 × 30 × 1500 mm aluminum alloy frames and trolleys were used to fix the vision sensor at a distance of 3.2 m from the ground to collect depth images of top-view of dairy cow, as shown in Fig. 2.

Collection process is carried out at two feeding times for dairy cows at 11:00 and 16:30. Dairy cows are fixed by neck clamps while eating. Collectors push the trolley mobile device behind cows to collect image data on dairy cows in a cowshed. At the same time, trained scorers are required to perform artificial dairy cow body condition scoring while collecting data. In order to reduce impact of the subjectivity of manual scoring, this experiment arranged two trained scorers to evaluate individual cow’s body condition score during data collection process, and take the average of the two as final body condition scoring result.

2.2. Data processing

2.2.1. Noise point removal

As Fig. 3 shows, it can be seen that the point cloud contains plenty of noise points, which should be removed. Since the noise points are more scattered than dairy cow area, the radius outlier removal method, which is also called point cloud filtering based on connectivity analysis, is used to remove noise points. The filtering operation is performed by setting the number of points contained in a given radius neighborhood as a threshold (Popescu and Lungu, 2014). The principle of radius filtering is shown in Fig. 4(a), set the radius of the search point to $\boldsymbol{r}$, the number of near-neighbor points is 4, only point $\boldsymbol{a}$ meet the conditions, so remove point $\boldsymbol{b}$ and point $\boldsymbol{c}$. The filtering results is shown in Fig. 4(b), it can be seen that the noise above point cloud is basically removed.

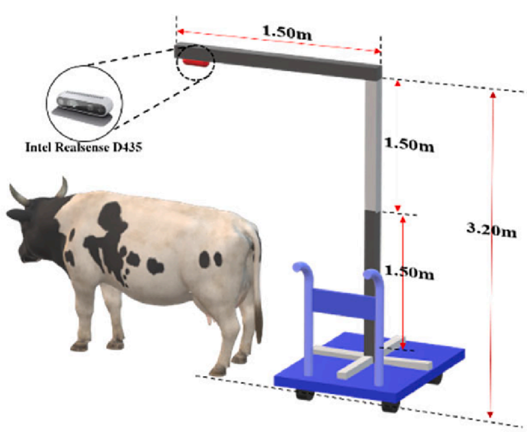


Fig. 2. Data Collection System.

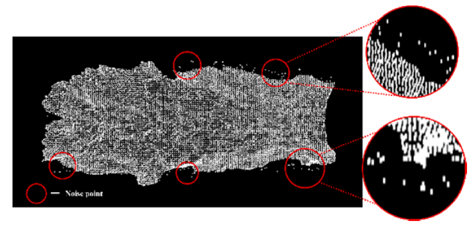


Fig. 3. Point Cloud Generated from Depth Image.

2.2.2. Region of interest extraction

Since most local area features used as a reference in body condition score are concentrated in dairy cows’ back end, in order to reduce the influence of redundant features in other regions, dairy cows’ back end area is chosen as region of interest and extracted. The back end and the other region of dairy cow have the lumbar horn bone connection as obvious boundary, as shown in Fig. 5(a), and the back end area has been marked using scalar fields with RGB chart, the other part was marked using scalar fields with grey chart. While two lumbar horn bones have the largest span on X-axis. Therefore, traverse point cloud data in turn, after obtaining the points corresponding to the maximum and minimum values in X-axis direction, the front part of dairy cow is removed using the line between two points to obtain dairy cows’ back end as region of interest, as shown in Fig. 5(b).

2.2.3. Region of interest processing

Due to different heights of different cows, the coordinate values of points of dairy cows with same score are quite different, which affects extraction of features by neural network. Therefore, point cloud of dairy cattle is translated according to the minimum value of point cloud XYZ, as shown in Fig. 6, (a) represents the point cloud to be translated, (b) represents the translation result.

At the same time, in order to sample uniformly and reduce computational load, this experiment uses voxel grid filtering to down-sample ROI point cloud, the principle of this method is as follows:

1.Obtain the maximum and minimum values of X, Y, Z axes in point cloud and obtain the length L, width W and height H of the box enclosing the point cloud. Then set voxel grid edge length to vl, and use the grid to divide enclosing box into I, J, K voxel grids, as shown in equation (1).

$$\left\{\begin{array}{l} I = |L/vl| \\ J = |J/vl| \\ K = |K/vl| \end{array}\right. \tag{1}$$

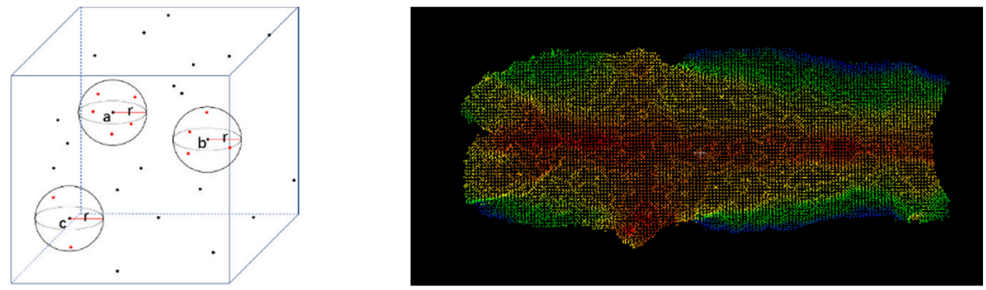
2.Use (α, β, γ) to number each voxel grid, as shown in equation (2).

$$\left\{\begin{array}{l} \alpha = |(x_\alpha - x_{min})/vl| \\ \beta = |(y_\beta - y_{min})/vl| \\ \gamma = |(z_\gamma - z_{min})/vl| \end{array}\right. \tag{2}$$

3.Finally, refining point cloud, calculating the center of gravity of all points $G_{\alpha,\ \beta,\ \gamma}$ in each voxel grid, and replace all points in the voxel grid with the center of gravity, as shown in equation (3), of which n represented the point number of point cloud, p_m represented each point in point cloud (Rusu and Cousins, 2011). , the down-sampling result is shown in Fig. 7.

$$G_{\alpha,\beta,\gamma} = \frac{1}{n}\sum_{m=1}^{n} p_m \tag{3}$$

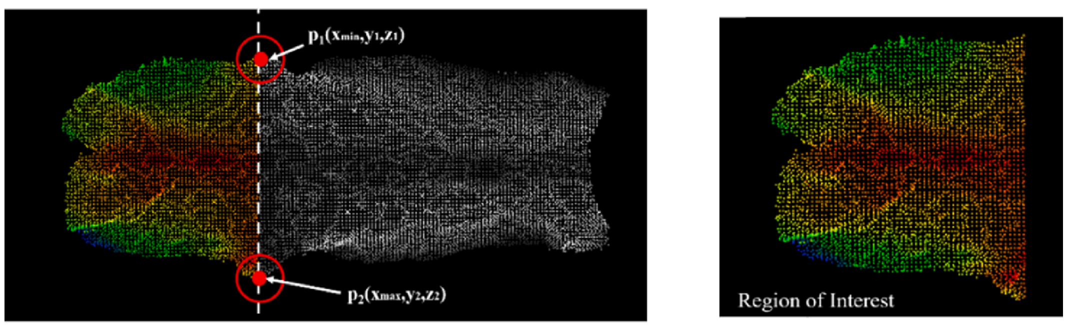
In addition, because the concavity features are mainly reflected by the Z-axis coordinate value, in order to highlight the concavity degree of the region of interest on dairy cow, the Z-axis coordinate value is enlarged by 1.25 times to obtain more obvious convex-concave features. Due to the feature could be extracted of the point cloud translated from the depth image are only the X, Y, Z three axis coordinate values. The surface normal is an important feature of geometric surfaces. In order to facilitate network to extract more morphological features, this work



(a) Principle of Outlier Removal Filtering

(b) Filtering result

Fig. 4. Radius Outlier Removal Filtering Principle and Result.



(a) Region of interest locating

(b) Region of interest

Fig. 5. Region of interest extraction.

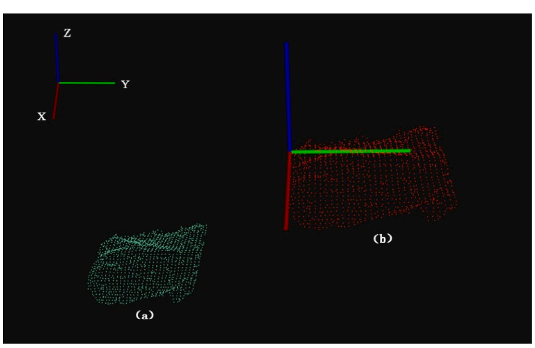


Fig. 6. Translation Operation Result.

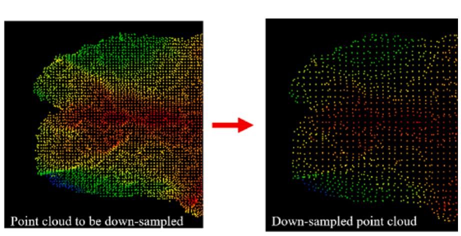


Fig. 7. Process of down-sampling.

adds the X, Y, Z three-direction normal vectors to features of each point. This article uses local surface fitting to estimate normal vector. This method calculates a local plane XOY in the least square sense based on k nearest neighbors of a certain point. The normal P_i of XOY can be regarded as the normal of this point (Hoppe et al., 1992), and calculation process of normal vector is shown in Fig. 8.

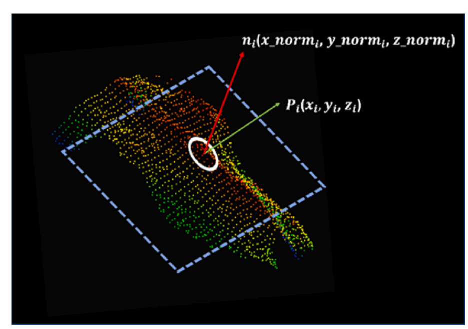


Fig. 8. Normal vector calculation process.

The feature vector of each point in final point cloud is shown in equation (4). Finally, the processed point cloud data is constructed according to different body condition scores to construct a data set.

$$P_i = \left(x_i, y_i, z_i, x_{normi}, y_{normi}, z_{normi}\right)^T, i = 1, 2, \cdots, N \tag{4}$$

2.3. Point cloud feature extraction with attention-guided

Due to BCS estimation mostly rely on the concavity degree of local area of ROI, so body condition scoring model should have better extraction capabilities for local features. In addition, plenty of redundant information exists in region of interest. Therefore, in order to extract important features, attention guiding is introduced. In order to better extract the feature of the point cloud data of dairy cow, and then construct automatic body condition scoring model, a point cloud feature extraction network with attention-guided was proposed in this research. PointNet++ (Qi et al., 2017a), a point cloud feature extraction network with hierarchical structure, was selected as basic network, and attention guidance is introduced to improve the extraction capabilities of local important features, the structure of the attention guided point cloud feature extraction network is shown in Fig. 9. Firstly, PointNet++ is a classic point cloud feature extraction neural network that are often used in classification and segmentation of point cloud object. Especially, PointNet++ introduced hierarchical structure, which make network provide a higher level of feature on increasingly large areas. In term of the structure of network, it consists of a series of set abstraction levels, including Sampling Layer, Grouping Layer and PointNet layer.

Sampling layer uses the farthest point sampling method to select the center point of each region, which can ensure the uniform sampling of the sample, the implementation of farthest point sampling method is as follows: 1)For an input point cloud with N points, point P_0 is selected as the initial point to build sampling set $S = \{P_0\}$; 2) y calculating the distance from the points other than P_0 to point P_0, the point with the maximum distance is selected as P_1 to join the sampling set S, $S = \{P_0, P_1\}$; 3) Then calculate the distance between remaining points and the points in the point set, and take the minimum value as the distance from that point to the point set. After calculating the distance from all points to the point set, take the point P_2 with the maximum distance and add it to the sampling set, $S = \{P_0, P_1, P_2\}$; 4) Repeat this step until the target number of points is sampled. After the center point is selected, the Grouping layer uses the ball query to get the set of region points centered on that point.

PointNet layer uses PointNet network to extract features from the region of interest. Due to disorder and rotation of point cloud data, traditional convolution method of 2D images is not applicable to point cloud data. To solve the problems caused by two characteristics of point clouds, PointNet uses Spacial Transform Network (STN) to learn position information of point clouds themselves to obtain a rotation matrix that facilitates the classification to solve rotation problem, and uses Max-Pooling symmetry function to solve disorder problem. PointNet extracts features mainly by multi-layer perceptron with MaxPooling.

First, point cloud is input to a T-Net network for spatial transformation to solve rotation problem, and output point cloud is used to raise point cloud feature dimension to 64 by using Multi-Layer Perceptron (MLP, Multi-Layer Perceptron), and then a spatial transformation with two MLP layers to 1024, and Max Pooling is used to extract global features, the structure of PointNet is shown in Fig. 10.

Attention mechanism can be used to screen important information from a large amount of information and focused on important information while training. In computer visual field, attention mechanism has different forms of implementation (Vaswani et al., 2017), and existing attention mechanism method can be roughly divided into two: Soft Attention and Hard Attention (Zhao et al., 2017). Hard attention specifies whether the part of input data is concerned, pay attention to given weight of 1, otherwise 0; Soft Attention was used in this study, which is essentially a continuous distribution problem between [0, 1], depending on the degree of attention, different weight values are given to [0, 1] interval. Calculation process of attention mechanism can be divided into two steps. First, weight coefficient calculation is performed, followed by weight based on weight coefficient to value, where first step can be divided into two phases, similarity calculation and normalization. In this study, calculating the similarity between two points by a dot product method first, if the result of dot product grows large in magnitude, softmax function result will be pushed into regions where it has extremely small gradients (Gao et al., 2021). To offset this effect, this study multiplies the result of dot product calculation by a scaling factor *1/L*. Implementation process is as shown in equation (5).

$$Similarity(p_i, p_j) = p_i\hat{A}\cdot p_j/\mathrm{L}\ (\mathrm{j} = 1, 2, \cdots, \mathrm{N}) \tag{5}$$

Among them, p_i and p_j are points in point cloud (j = 1, 2, …, N), and L is scaling factor. After obtaining the similarity between pi and p_j (i, j = 1, 2, …, N), use the softmax function to transform it into a probability distribution, that is, a weight value. Finally, weight value is multiplied with input, as shown in equation (6), as shown in Fig. 11.

$$Attention\left(P_i^{'}\right) = \sum_{j=1}^{N} Softmax\left(Similarity\left(P_i, P_j\right)/L\right)\hat{A}\cdot P_j \tag{6}$$

2.4. Body condition score estimation

Combined with the above problems and the experimental data of this study, the estimation of body condition score can be regarded as a multi-classification problem. For the experimental data of this study, body condition score of dairy cow as study object was 2.0 to 4.0, with an increment of 0.25, therefore, the number of categories for the classification task done in this study is 9. After using PointNet++ with attention guiding to extract feature of region of interest, and there was a 9-dimensional softmax layer. Softmax is used in the multi-classification process, which maps the output of multiple neurons into the (0,1) interval, which can be understood as probabilities, to perform multi-classification. In the final selection of the output node, we can then select the node with the highest probability as our prediction target, then get the body condition score.

3. Result and discussion

3.1. Development tools

The experiment was conducted on a hardware configuration with Intel Core i7-7800 × 3.5 GHz CPU, 16 GB memory and NVIDIA GeForce RTX 2060 GPU with 6 GB memory.

3.2. Evaluation metrics

Based on the performance of generative model on test set, this paper

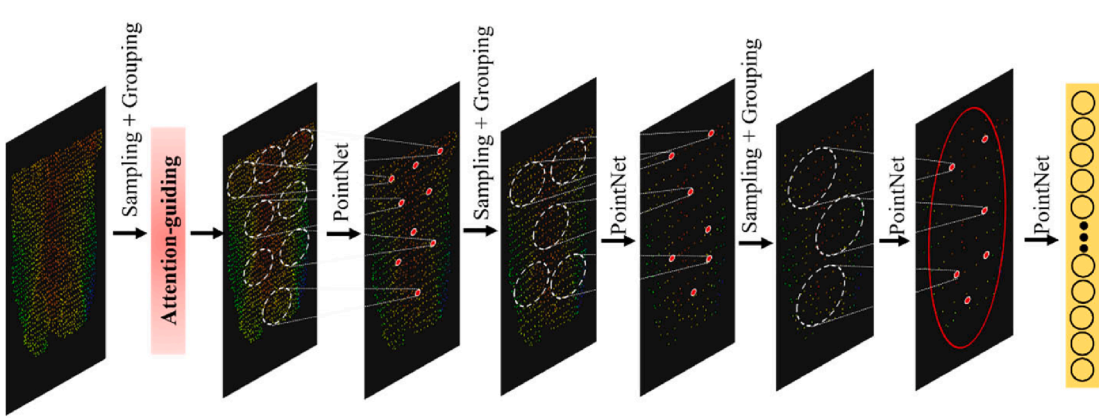


Fig. 9. Structure of point cloud feature extraction network with attention guided.

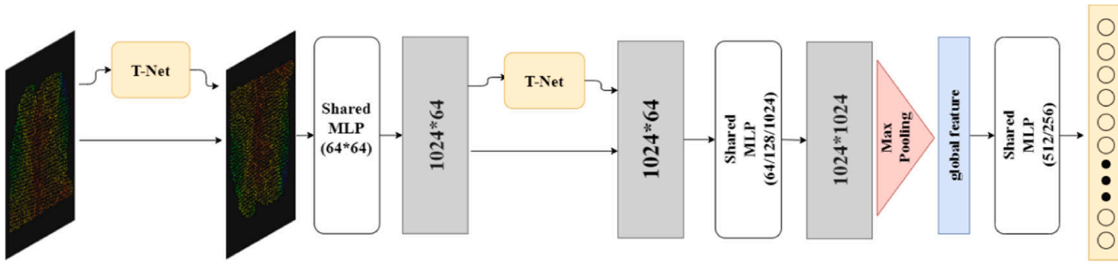


Fig. 10. The structure of PointNet.

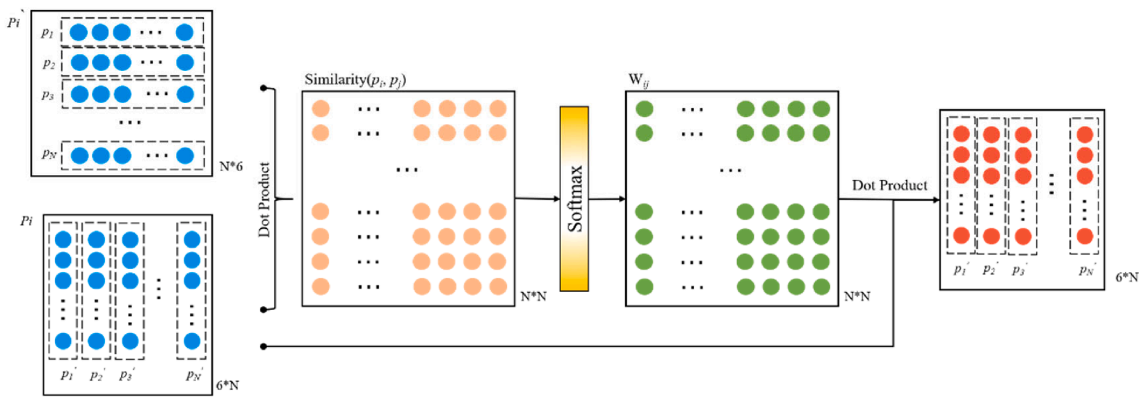


Fig. 11. Attention calculation process.

calculates a set of indicators to evaluate BCS model. In multi-classification tasks, confusion matrix is often used to measure and analyze model performance. The vertical axis of confusion matrix represents predicted value, and horizontal axis represents true value. Classification results of each category can be calculated as follows: the amount of truly classified samples (True Positive, TP), the amount of truly classified samples which don't belong to the true class (True Negative, TN), the number of samples incorrectly classified (False Positive, FP) or not recognized as correct class (False Negative, FN). The calculation method of the above classification task evaluation metrics is shown in the Table 1.

Based on above values, the following metrics used to evaluate BCS model are calculated:

1. Precision: Ratio of the number of correctly identified samples to all results identified as sample, used to evaluate the ability of sub-model to not predict a negative sample as a positive sample.

$$Precision = \frac{TP}{(TP + FP)}$$

2. Recall: Ratio of the number of correctly identified samples to total number of samples of this type, and the ability of evaluation model to find all positive samples.

$$Recall = \frac{TP}{(TP + FN)}$$

Table 1
Accuracy of body condition scoring model based on 3D and 2D features.

Experiment	Accuracy under different point deviation		
	0	0.25	0.50
BCS model based on 2D feature	0.31	0.52	0.63
BCS model based on 3D feature	0.41	0.70	0.90

3. F1-score (Maimon and Rokach, 2010): Harmonic average of Precision and Recall, used to measure the effect of classifier in presence of rare categories.

$$F1 - score = \frac{2*Precision*Recall}{(Precision + Recall)}$$

3.3. Comparison of body condition scoring based on depth image and point cloud

To compare the representation and extraction of 3D features in 2D digital images and 3D point cloud images. MobileNetV2 was used to extract feature and only depth image was used to construct BCS model. Then, point cloud data containing only the X, Y, Z axis information is used as input, PointNet++ was used to extract feature automatically. The performance in test set of the model generated from the two experiences are represented by the exact accuracy, accuracy with 0.25 and 0.50 BCS units, as shown in Table 1. It can be obviously seen that compare to the BCS model based on depth image, the model base on point cloud has better performance in exact accuracy, accuracy with 0.25 and 0.50 BCS units, which respectively improve 0.10, 0.18, 0.27.

The reason why the model constructed by depth information extraction network doesn't perform well in test can be concluded as follows: 1) As shown in Fig. 12, for ease of observation, convert depth image to pseudo-color image based on depth information. It can be seen that there are lots of noise in depth image, which can't be removed as simply as point cloud. 2) From the depth image show above, it only represents depth information through 256 depth levels, since the depth level Z and location X-Y are quantified, so depth image don't have the ability to accurately represent empty points. In addition, the depth information contained in image is not sequential, which could cause some of the depth information to be missing. That is the reason why this study selected point cloud as research object, which can represent the 3D location information better.

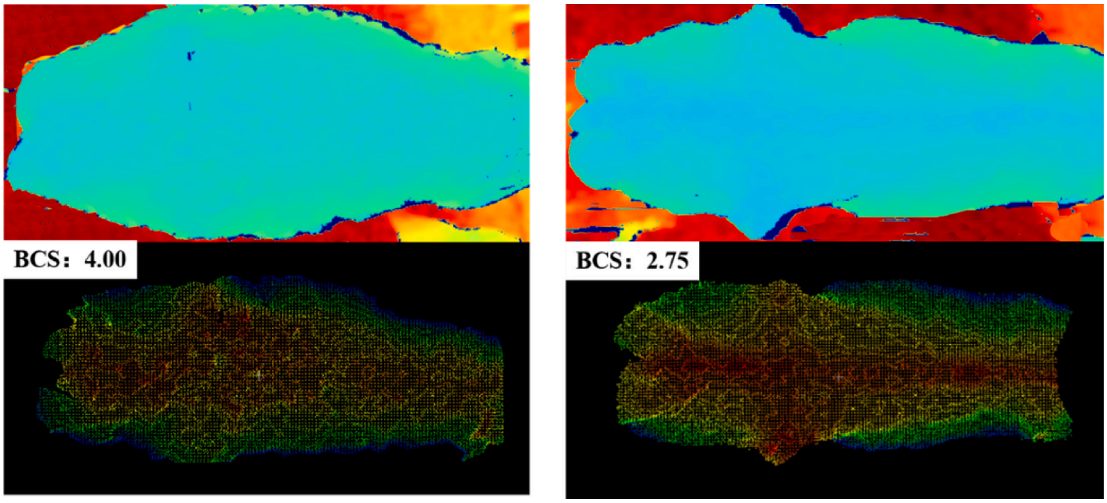


Fig. 12. Point cloud and depth image on the back of cows with different BCS.

3.4. Comparison of effect of different feature extraction network

Next, the effects of the experiments which used point cloud data as research object will be evaluated using confusion matrices. In machine learning, the confusion matrix is an error matrix that is often used to evaluate the performance of supervision learning algorithm. In

	2.00	2.25	2.50	2.75	3.00	3.25	3.50	3.75	4.00
2.00	5	3	2	1	3	0	0	0	0
2.25	7	22	19	22	12	5	2	0	0
2.50	4	6	17	8	16	5	3	2	0
2.75	4	17	20	89	34	25	6	3	0
3.00	1	13	28	33	93	27	15	4	1
3.25	0	4	5	9	11	18	5	3	1
3.50	0	2	1	4	17	9	13	4	2
3.75	0	1	1	4	4	2	3	5	3
4.00	0	0	1	1	5	3	4	3	7

(a)Experiment I

	2.00	2.25	2.50	2.75	3.00	3.25	3.50	3.75	4.00
2.00	6	4	1	2	1	0	0	0	0
2.25	9	26	19	23	9	2	0	1	0
2.50	2	6	22	11	14	4	2	0	0
2.75	3	21	24	95	35	17	1	2	0
3.00	1	9	24	29	114	19	15	3	1
3.25	0	3	2	6	15	22	4	3	1
3.50	0	2	1	5	16	9	14	4	1
3.75	0	0	1	2	3	3	4	7	3
4.00	0	0	0	2	4	3	4	3	8

(b) Experiment II

	2.00	2.25	2.50	2.75	3.00	3.25	3.50	3.75	4.00
2.00	6	4	2	1	1	0	0	0	0
2.25	14	29	22	20	3	1	0	0	0
2.50	4	13	25	11	6	1	1	0	0
2.75	1	19	21	97	38	20	0	2	0
3.00	0	6	18	37	105	25	20	3	1
3.25	0	2	1	6	15	24	5	3	0
3.50	0	0	1	5	9	13	16	7	1
3.75	0	0	0	1	3	3	2	10	4
4.00	0	0	0	1	2	4	3	5	9

(c)Experiment III

	2.00	2.25	2.50	2.75	3.00	3.25	3.50	3.75	4.00
2.00	8	4	1	1	0	0	0	0	0
2.25	16	42	16	12	3	0	0	0	0
2.50	3	9	27	12	7	1	2	0	0
2.75	0	24	23	100	32	18	1	0	0
3.00	0	5	12	36	117	26	17	2	0
3.25	0	1	1	6	15	25	4	4	0
3.50	0	0	1	2	7	9	19	10	4
3.75	0	0	0	1	0	2	7	12	1
4.00	0	0	0	1	2	1	4	7	9

(d) Experiment IV

Fig. 13. Confusion Matrix of Different Experiment. (a) Experiment I: Coordinate information as feature input, PointNet was used to extract feature; (b) Experiment II: Coordinate information as feature input, PointNet++ was used to extract feature; (c) Experiment III: Coordinate information and normal vector as feature input, PointNet++ was used to extract feature; (d) Experiment IV: Coordinate information and normal vector as feature input, PointNet++ with attention guiding was used to extract feature.

confusion matrix, value on main diagonal is accurately predicted. In ideal state, numerical value in matrix should be all distributed on main diagonal, but the results of confusion matrix of these experiments are not achieved. However, from the whole, most predictive results are concentrated near main diagonal, while due to difference in feature of the increment of 0.25 is not obvious, even experienced scorers are also complicated to distinguish, and the amount of dairy data in this range is large.

Firstly, two BCS models constructed by PointNet++ and PointNet. The performance in the test set of these two models generated are shown in Fig. 13(a) and Fig. 13(b) respectively, the accuracy of the model calculated by the confusion matrix under different manual errors is shown in Table 2.

Through two confusion matrices and Table 2, the exact accuracy, accuracy with 0.25 and 0.50 BCS units of experiment I are 0.37, 0.64, 0.86, respectively. While exact accuracy, accuracy with 0.25 and 0.50 BCS units of experiment II are 0.41, 0.70, 0.90, respectively. That is because PointNet uses maxpooling on global region of interest, which could reduce the ability of the perception of local area feature. While the ability to extract local feature of PointNet++ is enhanced by introducing a hierarchical structure, through the sampling layer, the grouping layer divides the points in the region of interest into local regions and uses PointNet for feature extraction in the local regions. Therefore, it is more suitable for the construction of automatic scoring model of dairy body condition.

3.5. Comparison of BCS model with or without normal vector feature

Considering that the performance of body condition scoring model base on PointNet++ with only take XYZ coordinate value as input, which can extract fewer feature. Therefore, to let network extract more feature information, in basis of original features, this study calculated normal vector of each point used the method of partial surface fitting and construct BCS model. After adding normal vector, model is used to evaluate model in test set, classification result is shown in Fig. 13(c). As Table 2 shows, the exact accuracy, accuracy with 0.25 and 0.50 BCS units of experiment III are 0.45, 0.77, 0.94, respectively. The classification accuracy is improved 0.04, 0.07and 0.04 compared to model that only uses XYZ coordinate value as input.

Due to the surface normal is an important attribute of geometry surface, and features only extracted by location information is fewer, the introduction of normal vector can let the neural network extract more features, which could improve the effect of BCS model through that.

3.6. Comparison of BCS model with or without attention guided

In order to verify whether attention guided operation improves the network in perception of local features, this study took point cloud data, which have added normal vector, as input, using PointNet++ network to construct a body condition automatic scoring model, the training loss curve is shown in Fig. 14, comparing the performance of model without attention guided in test set. As shown in Fig. 9 (d), it is a confusion matrix of body condition automatic scoring model constructed by PointNet++ with attention guiding. As Table 2 shows, the exact accuracy, accuracy with 0.25 and 0.50 BCS units of experiment IV are 0.49, 0.80, 0.96, respectively. Compared with model without attention guiding, exact accuracy, accuracy of 0.25 and 0.50 errors are improved 0.04,0.03,0.02, respectively. Fig. 14 compares the performance of the 4 models for exact prediction, 0.25 and 0.50 manual error, the accuracy has obviously improved.

Table 2
Accuracy under different artificial errors of different experiment.

Experiment	Accuracy under different point deviation		
	0	0.25	0.50
Experiment I	0.37	0.64	0.86
Experiment II	0.41	0.70	0.90
Experiment III	0.45	0.77	0.94
Experiment IV	0.49	0.80	0.96

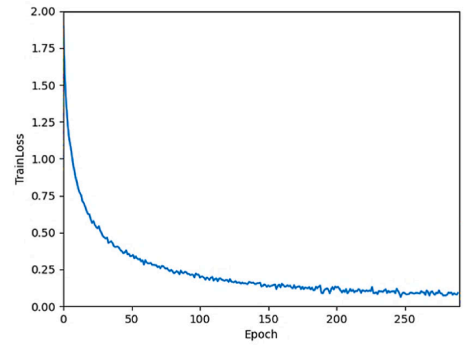


Fig. 14. Loss curve of the training process.

By using the attention mechanism to guide the neural network in the extraction of point cloud features, it is possible to improve the performance of the scoring model by enhancing the perception of local areas and further enhancing the perception and extraction of important features during the training process.

Table 3 is precision rate, recall rate and F1-score of 0, 0.25 and 0.50 BCS units of body condition scoring model base on PointNet++ with attention guiding.

3.7. Comparison to related works

Finally, automatic scoring model of experiment obtained by this study was compared with other scholars. Table 4 shows the comparison to related works and this study within different error range. It can be seen that using point cloud data as research objects is feasible for the construction of automatic body condition scoring model for dairy cow. Moreover, more abundant point cloud features and attention guiding can also improve perception of network on local features, as well as the effects of classification.

4. Conclusion

In order to automatically extract body condition score related feature from 3D data directly and automatically, then estimate the body condition of dairy cow. This paper proposes an automatic scoring method for dairy cow body condition based on attention-guided point cloud feature extraction. First, considered to learn 3D shape features, a representation of 3D data, which is close to raw sensor data and can represent 3D shape more sparsely and accurately, was used. After a series of data processing operations, which contained noise point removal, ROI extraction and translation, down-sampling and normal vector added, a feature extraction network based on point cloud was introduced. Moreover, a vision attention mechanism was designed and integrated to the network to guide the network to learn the concavity information more significantly. The estimation accuracy of the body condition estimation model based on feature extraction network with attention guided is 80% within 0.25 error and 96% within 0.5 error. Compared with previous work, point cloud classification network with attention guiding is used in comparable results have been achieved in accuracy of BCS estimation.

CRediT authorship contribution statement

Wei Shi: Conceptualization, Methodology, Investigation, Data curation, Writing – original draft, Writing – review & editing. **Baisheng Dai:** Conceptualization, Methodology, Funding acquisition, Software,

Table 3
Evaluation indicators on test set.

BCS	exact predictions			0.25 range error			0.50 range error		
	Precision	Recall	F1	Precision	Recall	F1	Precision	Recall	F1
2.00	0.30	0.57	0.39	0.80	0.86	0.83	1.00	0.93	0.96
2.25	0.49	0.47	0.48	0.71	0.83	0.77	0.93	0.97	0.95
2.50	0.33	0.44	0.38	0.76	0.79	0.77	0.97	0.95	0.96
2.75	0.58	0.51	0.54	0.87	0.78	0.82	0.98	0.99	0.99
3.00	0.64	0.54	0.59	0.90	0.83	0.87	0.98	0.97	0.97
3.25	0.30	0.45	0.36	0.67	0.79	0.72	0.96	0.96	0.96
3.50	0.35	0.37	0.36	0.61	0.73	0.67	0.94	0.94	0.94
3.75	0.34	0.52	0.41	0.77	0.87	0.82	0.92	0.96	0.94
4.00	0.64	0.38	0.47	0.80	0.67	0.73	1.00	0.83	0.91
Average	0.44	0.47	0.44	0.77	0.79	0.78	0.96	0.94	0.95

Table 4
Comparison to related works and this study within different point deviation.

Error range	Anglart (2010)	Spoliansky (2016)	Alvarez (2018)	Liu (2020)	Our Method
0.00	–	0.40	–	–	**0.49**
0.25	0.69	0.74	0.78	0.76	**0.80**
0.50	0.95	0.91	0.94	0.94	**0.96**

Validation, Formal analysis, Writing – review & editing. **Weizheng Shen:** Methodology, Funding acquisition, Data curation, Writing – review & editing, Supervision. **Yukun Sun:** Resources, Data curation, Writing – review & editing. **Kaixuan Zhao:** Methodology, Resources, Validation. **Yonggen Zhang:** Validation, Data curation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Acknowledgements

This work was supported in part by the National Natural Science Foundation of China (grant number 32072788, 31902210 and 32002227), in part by the National Key Research and Development Program of China (grant number 2019YFE0125600), and in part by the earmarked fund for CARS (grant number CARS-36).

References

Alvarez, J.R., Arroqui, M., Mangudo, P., et al., 2018. Body condition estimation on cows from depth images using Convolutional Neural Networks. Comput. Electron. Agric. 155, 12–22. https://doi.org/10.1016/j.compag.2018.09.039.

Anglart, D., 2010. Automatic estimation of body weight and body condition score in dairy cows using 3D imaging technique (Master's thesis). Faculty of Veterinary Medicine and Animal Science, Swedish University of Agricultural Sciences.

Azzaro, G., Caccamo, M., Ferguson, J.D., et al., 2011. Objective estimation of body condition score by modeling cow body shape from digital images. J. Dairy Sci. 94 (4), 2126–2137. https://doi.org/10.3168/jds.2010-3467.

Bercovich, A., Edan, Y., Alchanatis, V., et al., 2013. Development of an automatic cow body condition scoring using body shape signature and fourier descriptors. J. Dairy Sci. 96 (12), 8047–8059. https://doi.org/10.3168/jds.2013-6568.

Bewley, J., Peacock, A., Lewis, O., et al., 2008. Potential for estimation of body condition scores in dairy cattle from digital images. J. Dairy Sci. 91 (9), 3439–3453. https://doi.org/10.3168/jds.2007-0836.

Zhao, B., Wu, X., Feng, J., et al., 2017. Diversified visual attention networks for fine-grained object classification. IEEE Trans. Multimedia. 19 (6), 1245–1256. https://doi.org/10.1109/TMM.2017.2648498.

Zin, T.T., Seint, P.T., Tin, P., et al.. Body Condition Score Assessment of Depth Image using Artificial Neural Network. https://doi.org/10.1145/3408066.3408103.

Qi, C.R., Yi, L., Su, H., et al., 2017. PointNet++: Deep hierarchical feature learning on point sets in a metric space. 31st International Conference on Neural Information Processing Systems (NIPS), pp. 5105–5114.

Ferguson, J.D., Galligan, D.T., Thomsen, N., 1994. Principal descriptors of body condition score in holstein cows. J. Dairy Sci. 77 (9), 2695–2703. https://doi.org/10.3168/jds.S0022-0302(94)77212-X.

Gao, X.Y., Wang, Y.Z., Zhang, C.X., et al., 2021. Multi-head self-attention for 3D point cloud classification. IEEE Access. 9, 18137–18147. https://doi.org/10.1109/ACCESS.2021.3050488.

Halachmi, I., Klopčič, M., Polak, P., et al., 2013. Automatic assessment of dairy cattle body condition score using thermal imaging. Comput. Electron. Agric. 99 (6), 35–40. https://doi.org/10.1016/j.compag.2013.08.012.

Liu, D., He, D., Norton, T., 2020. Automatic estimation of dairy cattle body condition score from depth image using ensemble model. Biosyst. Eng. 194, 16–27. https://doi.org/10.1016/j.biosystemseng.2020.03.011.

Rusu, R.B., Cousins, S.. 3D is here: Point Cloud Library (PCL). https://doi.org/10.1109/ICRA.2011.5980567.

Hoppe, H., DeRose, T. Duchamp, T., et al. Surface reconstruction from unorganized points. ACM SIGGRAPH Computer Graphics. 1992, 26(2), 71-78. doi: 10.1145/142920.134011.

Maimon, O., Rokach, L., 2010. Data mining and knowledge discovery handbook, 2nd ed. Springer Publishing Company, Incorporated. https://doi.org/10.1007/978-0-387-09823-4_42.

Popescu, C.R., Lungu, A., 2014. Real-Time 3D reconstruction using a Kinect sensor. Computer Science & Information Technology 2 (2), 95–99. https://doi.org/10.13189/csit.2014.020206.

Qi, C. R., Su, H., Mo, K., et al., 2017. PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 77-85. https://doi.org/10.1109/CVPR.2017.16.

Spoliansky, R., Edan, Y., Parmet, Y., et al., 2016. Development of automatic body condition scoring using a low-cost 3-dimensional Kinect camera. J. Dairy Sci. 99 (9), 7714–7725. https://doi.org/10.3168/jds.2015-10607.

Vaswani, A., Shazeer, N., Parmar, N., et al. (2017). Attention is all you need. arXiv preprint arXiv:1706.03762.

Sun, Y., Huo, P., Wang, Y., et al., 2019. Automatic monitoring system for individual dairy cows based on a deep learning framework that provides identification via body parts and estimation of body condition score. J. Dairy Sci. 102 (11) https://doi.org/10.3168/jds.2018-16164.

Wildman, E., Jones, G., Wagner, P., et al., 1982. A dairy cow body condition scoring system and its relationship to selected production characteristics. J. Dairy Sci. 65 (3), 495–501. https://doi.org/10.3168/jds.S0022-0302(82)82223-6.