IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 3 December 2023, accepted 29 December 2023, date of publication 2 January 2024,
date of current version 11 January 2024.

Digital Object Identifier 10.1109/ACCESS.2023.3349320

RESEARCH ARTICLE

Body Condition Scoring of Dairy Cows Based on Feature Point Location

KEQIANG LI[1,2,3], GUIFA TENG[4,5], JIANTAO WANG[6], YUXIN ZHANG[6], LEI GAO[7], AND HUI FENG[6]

[1]School of Mechanical and Electrical Engineering, Hebei Agricultural University, Baoding 071001, China
[2]Hebei Innovation Center for Smart Perception and Applied Technology of Agricultural Data, Qinhuangdao 066004, China
[3]School of Mathematics and Information Technology, Hebei Normal University of Science and Technology, Qinhuangdao 066004, China
[4]Hebei Digital Agriculture Industry Technology Research Institute, Hebei Agricultural University, Shijiazhuang 050081, China
[5]Hebei Key Laboratory of Agricultural Big Data, Baoding 071001, China
[6]Tangshan Animal Husbandry Technology Promotion Station, Tangshan 063030, China
[7]Qianan Bureau of Agriculture and Rural Affairs, Qianan 064400, China

Corresponding author: Guifa Teng (tguifa@126.com)

This work was supported in part by the National Natural Science Foundation of China under Grant U20A20180, and in part by the Open Fund Project of Hebei Agricultural Data Intelligent Perception and Application Technology Innovation Center under Grant ADIC2023Y007.

This work involved human subjects or animals in its research. The author(s) confirm(s) that all human/animal subject research procedures and protocols are exempt from review board approval.

ABSTRACT The use of computer vision to estimate the Body Condition Score for cow has demonstrated to be feasible. However, most research has been limited to fixed camera positions, which restricts the technique's usefulness. This research acquired cow data at various distances and angles to investigate the impact of distance and different depth images restoration method on scoring accuracy. A U-Net neural network-based model was employed for background segmentation. The model proposed in this study performs best among SegNet, UNet, UNet++ and DeeplabV3, with significant performance improvement and superior overall segmentation accuracy, and has stronger foreground object identification capability. Additionally, we utilize the concept of human feature point localization to pinpoint the positions of cow feature points. The results show that compared to Hourglass, CPN, and Hrnet, the model in this study has significant advantages in three core indicators: accuracy, recall, and mAP. Moreover, we presented an unsupervised depth image reconstruction model based on the Denoising Diffusion Probabilistic Model and Unet++ Mode, facilitating the measurement and scoring of cow characteristics. Finally, the measurement results of cow body size using the Lerp model, autoregressive model, GAN model, and the depth image completion model proposed in this study were compared. The method proposed in this study was found to be effective and feasible for meeting actual production requirements at a camera distance of 1-2 m from the cow, achieving high accuracy with a coefficient of determination above 0.9. Additionally, the three models used were effective in handling small-scale depth deficits, with a high degree of agreement between machine and manual measurements. The DDPM-based depth image reconstruction model was found to produce the most accurate results when the camera distance from the cow was between 1-3m. Therefore, this research makes it possible to make flexible measurements in the near range using existing methods. Accurate measurements at longer ranges are influenced by depth image quality and feature point localization accuracy, which require more advanced techniques for further investigation.

INDEX TERMS Body condition score, feature point location, denoising diffusion probabilistic model.

I. INTRODUCTION

Precision animal husbandry has enabled producers to gain valuable insights into the health and overall welfare of livestock through scientific feeding and precise management.

The associate editor coordinating the review of this manuscript and approving it for publication was Yudong Zhang.

In particular, the body condition score (BCS) of cows is an indicator of body energy reserves, and it can be used to assess whether cows are in appropriate condition at each stage of the lactation. This information can aid in evaluating the feeding management and ensuring that nutritional levels are appropriate, thereby predicting herd productivity and providing a high reference value for farm management and decision-making [1]. Traditional scoring methods for BCS rely on tactile or visual methods, which are subjective and prone to variations in interpretation among even well trained scorers [2], [3], [4], [5]. Moreover, the manual scoring processes is not only time-consuming and costly but also lead to personal injury and affect scoring accuracy due to the stress response of livestock during the measurement process.

In an attempt to overcome the limitations of traditional BCS measurement methods, researchers have explored alternative approaches involving 2D imaging. Bewley [6] manually selected feature points from images to score BCS, while Halachmi [7] predicted BCS by extracting the profile of cows and calculating the average absolute error between the fitted polynomial and the cow profile. Battiato [8] used statistical shape analysis and regression machines to evaluate BCS using top view images of cows, and found it is approach to be feasible. Based on these methods, Huang [9] proposed using the multi-box detector (SSD) method to evaluate cows by collecting 898 images of cow tails. The experiment showed that the classification accuracy of this approach for cow scoring was 98.46%.

Although there have been advancements in the methods for processing 2D images, the lack of 3D information due to the 2D projection limits further applications. However, in recent years, 3D sensors have emerged, providing depth information that can be useful for cow body condition scoring [10], [11]. Several studies have utilized 3D cameras to capture surface information of cows, extract feature points, and predict BCS. For instance, Fischer [12] selected four feature points manually, while Alvarez [13] used the Squeeze-Net model to predict BCS with high accuracy. Liu [14] proposed an image processing algorithm that can extract features automatically and has good performance in predicting extreme cow body condition scores. Martins [15] collected 13 feature points using the 3D camera and obtained the relationship between BCS and body weight. Overall, 3D images contain depth information that reflects the degree of body surface concavity and is more strongly correlated with cow body energy storage status. Therefore, 3D visual-based systems [5], [10], [12], [13], [14], [15], [16] tend to be more relevant than 2D-based systems [6], [7], [8] and have great potential for improve the accuracy of cow body condition scoring.

Ferguson [17] established the high correlation between cow body condition score and several anatomical regions, such as the thurl region, ischial and ileal tuberosities, iliosacral and ischiococcygeal ligaments, transverse processes, and spinous processes of the lumbar vertebrae. In practice, scorers determine the scoring range based on first impressions and then consider the different body features of cows in various regions to arrive at a result. However, current research has only partially quantified these features [11], [12], [13]. For instance, Spoliansky [10] used image processing techniques and regression algorithms to extract the cow tail data and automatically calculate the body condition score. Hansen [16] proposed a ''rolling ball'' algorithm to precisely extract the cow spine from depth images and detect cow weight, lameness, and body condition. Nonetheless, the quantification of cow body condition score features remains incomplete, and further research is necessary to understand the correlation between these features and cow body conditions score.

Prior studies have discussed the effects of using cameras at fixed distances for cow body condition scoring [18], [19], [20], but have failed to consider the impact of camera-to-cow distance on measurement accuracy. For example, Cozler [21] fixed five cameras to complete the morphological characterization of cows, while Zin [22] fixed a 3D camera at a distance of 1.8 meters from the back of the cow and achieved automated scoring of cows through image processing techniques and regression models. Li [23] estimated BCS by using a 3D surface fitting method with a camera mounted at a fixed distance of 2.4 m from the ground on a metal mount.

Furthermore, it is essential to acknowledge that the quality of depth images can be significantly impacted by variations in the distance between the 3D camera and the target, resulting in the presence of artifacts and voids. Previous research efforts have sought to enhance the quality of depth images through the application of depth map reconstruction methods [24], [25], [26]. However, these methods encounter a notable constraint: the absence of authentic depth images of actual livestock, rendering supervised learning approaches infeasible for training depth image complementation models. Consequently, the predominant approach in existing literature has been the utilization of filter-based methodologies to address gaps in depth maps. Nonetheless, it is important to note that filter-based techniques exhibit optimal efficacy in situations where the extent of missing depth values is limited. Once the missing depth region surpasses a certain threshold, the efficacy of filter-based depth image complementation method diminishes.

Due to the difficulty of acquiring ground truth in many production environments, unsupervised deep learning methods have received extensive attention in image reconstruction research. Unsupervised deep learning does not require labeled data for training. By learning the inherent feature representations within the data itself, it mines the hidden patterns behind the data distributions to obtain more generalizable feature representations. Compared with supervised learning, unsupervised deep learning reduces manual annotation workload, lowers over fitting risks, and achieves better feature transferability, expandability and computational efficiency [27], [28]. In recent years, researchers have proposed a remote sensing image super-resolution method called E2GAN, which designs modules to extract and enhance edge details to improve the edge and texture

quality of the generated images [29]. To improve computational efficiency and reduce network complexity, researchers proposed an efficient diffusion model based remote sensing image super-resolution method built upon the ideas of E2GAN, which achieves higher reconstruction quality by finely controlling the diffusion rate to reduce parameters [30].

It is noteworthy that the evaluation metrics used in these studies are mainly PSNR and SSIM, FID and visual subjective metrics. These metrics make judgments on the structural similarity between real images and generated images. The purpose of this study is to complement the missing depth values of corresponding positions in depth images for calculating cow body measurements. Therefore, evaluations based on structural similarity cannot determine the effectiveness of the method proposed in this study. Hence, the evaluation metrics used in these studies are not applicable to this research. Since no academic research has involved this aspect, there is still a lack of research on the impact of different image processing techniques on the depth values of cow keypoint features, thereby affecting the accuracy of scoring cow body conditions.

The aforementioned studies have demonstrated the theoretical foundation and operational simplification of using 3D cameras for cow body condition scoring. However, the practical application of this technology is limited by the insufficient quantification of features and the disregard for the impact of camera distance and depth image quality on the accuracy of scoring. To address these limitations, this study builds on prior research and proposed a feature point localization method for cow body condition scoring. Specifically, extract the cow's trunk from the background using the U-Net neural network model to enhance robustness against environmental factors. Then identify feature points based on the principles of human feature point localization, and a combination of color and depth images was utilized for scoring. Finally, we present an unsupervised depth image completion method based on Denoising Diffusion Probabilistic Model (DDPM) [31], and subsequently, we investigate the influence of distance and depth image quality on the process of assessing the body condition of dairy cattle.

II. MATERIALS AND METHODS

A. ETHICS STATEMENT

This experiment does not involve animal slaughter experiments. The data collection process of the relevant animals complies with the regulations of the Animal Ethics Committee of Hebei Agricultural University.

B. DATA ACQUISITION

The present investigation involved collecting video data of 150 adult cattle (Simmental and Holstein Friesian cattle, age >36 months of age) between April 2021 to August 2022. The data was collected using Intel RealSense D415 depth cameras (Intel™, Santa Clara, CA, USA) with a color horizontal field of view (±69.4°) and a depth horizontal field of view (±69.4°) at two different dairy farms(Luanzhou Jinshuo Dairy Farming Co. and Qian'an Erniu Livestock Farming Co). The cow was recorded while being gathered at the feed fences, while being photographed counter-clockwise around the cow from the highest point of the withers on the left side until the highest point of the withers on the right side. Notably, the quality of depth images was significantly impacted by the distance and this was taken into consideration during the collection process. During the collection process, the camera was positioned at distances of 1m-2m, 2m-3m, and 3m-4m from the cows, respectively, to ensure the feature points were visible in the view of the camera. Each cow was recorded for 30 seconds at a specific distance with a video frame rate of 30 frames per second and a resolution of 640 × 360 pixels for both the color and depth videos.

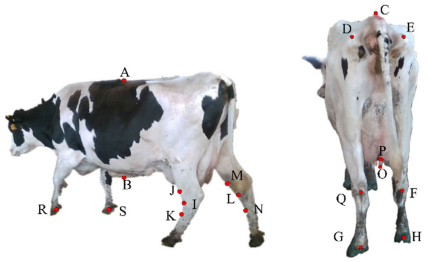


FIGURE 1. Schematic diagram of dairy cow feature point.

This research aimed to collect the body condition score of 150 cows using the 9-point scoring method of the Code of type classification in Chinese Holstein (GB/T35568-2017). Two trained technicians who had independently identified over 5000 cows were asked to evaluate 20 features using the aforementioned scoring method. In cases where there were discrepancies in the scores assigned by the two technicians, a third assessment was conducted to mitigate human measurement errors and obtain a final score.

This research investigation focused on analyzing a subset of features with high degrees of measure, namely chest depth, hip height, ischial width, parallelism of hind, hind leg curvature, length of teat, and depth of teat. In order to obtain the dataset of feature point localization for these features, manually screened 25200 images of cows captured by the camera positioned at various distances. The dataset was divided into training, validation, and test sets, with a 5:3:2 ratios being employed for this purpose.

C. DATA ANNOTATION

Experienced technicians used the LabeLimg annotation tool (version 1.8.6) to manually annotated color images. Figure 1 depicts the 19 feature point locations that were marked during the annotation process. Point A represented the lumbar point at maximum abdominal circumference. Point B was the lowest point of the abdomen at the maximum abdominal circumference. Point C was the highest point of the hip. Point D and E corresponded to the left and right ischial points respectively. Point F indicated the joint point of the left hind leg. Point G and H represented the left and right hind hooves

TABLE 1. Scoring criteria.

Scoring item \ Score	1	2	3	4	5	6	7	8	9
chest depth	<120:80	115:85	110:90	105:95	100:100	95:105	90:110	85:115	>80:120
hip height (cm)	<130	132	135	137	140	142	145	147	150
ischial width (cm)	≤10	12	14	16	18	20	22	24	≥26
parallelism of hind (°)	≥40	35	30	25	20	15	10	5	<5
hind leg curvature (°)	≥165	160	155	150	145	140	135	130	≤125
length of teat (cm)	≤2	3	3.5	4	5	6	7	8.5	≥10
depth of teat (cm)	≤-1	0	4	7	10	12	14	16	≥18

respectively. Point I, J, and K denoted the left leg hock point along with its upper and lower points. Point L, M, and N indicated the right leg hock point along with its upper and lower points. Point O and P were the bottom and top points of the teat respectively. Point Q referred to the joint point of the left hind leg. Point R and S represented the left and right front hooves respectively.

D. SCORING ITEMS AND CRITERIA

This study selected 7 out of 20 features based on their high degree of measure for measurement and analysis. The scoring features included chest depth, hip height, ischial width, parallelism of hind, hind leg curvature, length of teat and depth of teat.

(i) Chest depth Chest depth refers to the ratio of the vertical length from the lumbar spine to the bottom of the abdomen at the maximum abdominal circumference to the vertical length from the same point to the floor. This is equivalent to the ratio of the length from point A to point B to the length from point A to the plane composed of points R, S, G and H in Figure 1.

(ii) Hip height Hip height refers to the vertical length from the highest point of the hip bones to the floor. Correspond to the length from point C to the plane composed of points R, S, G and H in Figure 1.

(iii) Ischial width Ischial width refers to the length between ischial tubercles. Correspond to the length from point D to point E in Figure 1.

(iv) Parallelism of hind Parallelism of hind refers to the angle of the connection between the joint point of the left and right hind legs and the hind hoof. Correspond to the angle between the line QG and the line FG in Figure 1.

(v) Hind leg curvature Hind leg curvature refers to the Angle between the leg hock points and their upper and lower points. Correspond to the angle between the extension line JI and the extension line IK or the extension line ML and the extension line LN in Figure 1.

(vi) Length of teat Length of teat refers to the length between the Bottom and top of teat. Correspond to the length from point P to point O in Figure 1.

(vii) Depth of teat Depth of teat refers to the Relative length between the Bottom of teat and the hock. Correspond to the relative length from point P to line QF in Figure 1.

In accordance with the guidelines outlined in the Code of type classification in Chinese Holstein (GB/T35568-2017), whole numbers ranging from 1 to 9 were employed to reflect the extent of variance in cow body shape features from one extreme to another. Elaborated scoring criteria are depicted in Table 1.

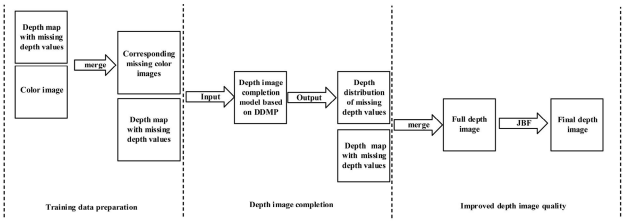


FIGURE 2. Depth image completion process.

E. IMAGE PROCESSING

The objective of the image processing procedure is to extract feature points from a color image and compute the corresponding score by incorporating the depth image. The workflow comprises of the subsequent steps: (i) restoration of the depth image by adopting distinct methods to minimize holes and texture-copying artifacts. (ii) Segment the body parts by combining the segmentation network from prior studies with data augmentation techniques, which allows automatic, remote and non-contact segmentation of the background. (iii) Localization and computation of feature points by employing HR-Net to augment the precision of feature point detection and produce the final scores.

1) DEPTH IMAGE RESTORATION

Depth cameras are capable of real-time depth information acquisition. However, depth images are susceptible to noise, missing regions, and other environmental factors, such as illumination and distance, which can significantly degrade their quality. To mitigate the impact of distance and illumination on the quality of depth images and enhance robustness in handling missing depth values, this study introduces an unsupervised depth image completion method based on the DDPM model. This method relocates the missing depth regions in depth images to their corresponding areas in color images, leveraging the prior information from the color images to progressively reconstruct the depth distribution within the damaged regions. The depth image completion process is illustrated in Figure 2.

The first phase involves the preparation of training data, as depicted in Figure 3. The training dataset comprises

incomplete depth images paired with corresponding color images containing corresponding imperfections. In practical agricultural settings, it is challenging to obtain depth images of livestock with complete depth information. Therefore, to maximize data utility and leverage the prior information from RGB images, this study saves the coordinates of depth value gaps in the depth images. Subsequently, based on these coordinates, the corresponding RGB information in the color images is removed, thus constituting the training dataset for the model. In this setup, incomplete depth images and their corresponding color images with position-based imperfections serve as inputs for paired training using the DDPM-based model.

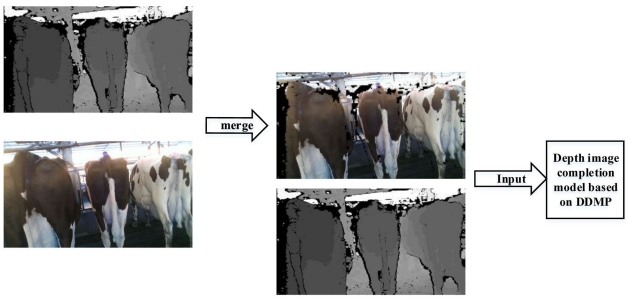


FIGURE 3. Training data preparation phase.

The second section outlines an unsupervised depth image reconstruction method based on the DDPM model. The DDPM Model is a parameterized Markov chain trained through variation inference. It diverges from the adversarial design of GANs by introducing the concept of time steps, enabling a smooth transition from simple to complex distributions. This gradual generation of samples from random noise aligns naturally with our depth image reconstruction task. The crux of DDPM lies in training a demising model. Given that noise and original data share the same dimensionality, we employ a U-Net model based on residual blocks and attention blocks for demising model training, as illustrated in Figure 4.

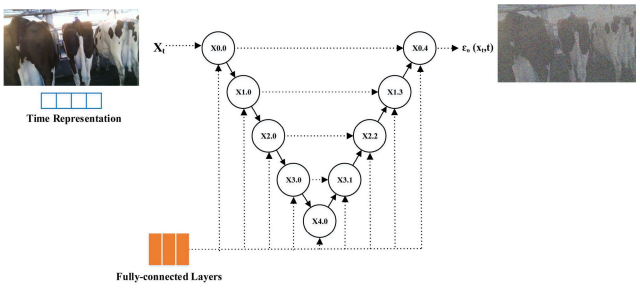


FIGURE 4. Noise prediction model U-Net in DDPM.

The U-Net architecture falls within the category of encoder-decoder frameworks. In this architecture, the encoder component is divided into different stages, each comprising down sampling modules to reduce the spatial dimensions of the features. Conversely, the decoder, as opposed to the encoder, gradually restores the features compressed by the encoder. Within the decoder module of U-Net, skip connections are introduced, involving the concatenation of features obtained from intermediate stages of the encoder. This inclusion of skip connections contributes to the optimization of the network. The encoding-decoding structure within U-Net aligns well with the concept of generating samples progressively from low-level to high-level features. By propagating multiscale information through skip connections, U-Net plays a pivotal role in refining structural information within RGB images.

In the U-Net structure of DDPM, the orange part of the Time Representation and Fully-connected Layers are utilized to provide information on time steps and feature integration respectively. Here, X_t denotes an intermediate state in the diffusion process, while (t) represents a specific time step. The Time Representation Layer ensures that the network can identify the current diffusion step and adjust its denoising behavior accordingly, and the Fully-connected Layers are responsible for processing and transmitting data and features within the network structure. Additionally, the formidable fitting capacity of U-Net and its variants empowers the DDPM to generate detailed and high-quality samples in an unsupervised context. This quality makes it particularly well-suited for recovering high-resolution depth images.

To further extract the correspondence between features in depth and color images, enhance the quality of depth image completion, and concurrently reduce model complexity, this study adopts the Unet++ model as a replacement for the Unet model used in DDPM for denoising purposes. The Unet++ model is illustrated in Figure 5.

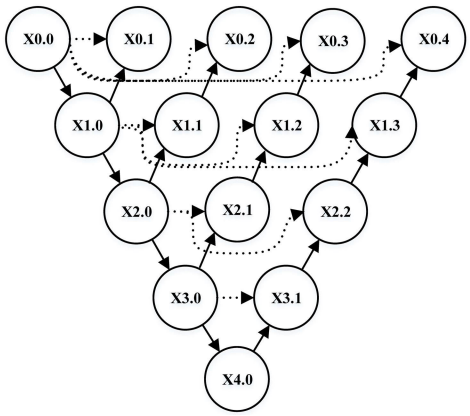


FIGURE 5. U-Net++ model.

The UNet++ architecture incorporates a deeper encoding-decoding structure with an increased number of skip connections. This structural enhancement facilitates the capture of global image information, contributing to effective denoising. Furthermore, UNet++ employs recursive dense connection modules, enabling the extraction of richer feature representations and enhancing the network's expressive capacity. Additionally, UNet++ leverages advanced training techniques such as residual learning and deep supervision, further augmenting the efficacy of network training. In comparison to the

conventional Unet model, UNet++ employs more advanced model compression and pruning techniques. It selectively removes less relevant structures and parameters for denoising tasks and fine-tunes training parameters to prevent parameter inflation caused by over fitting.

The third phase corresponds to the stage of improving depth image quality. Due to limitations in the number of images available in the training dataset, the depth image outputted in the second phase may still exhibit small areas of depth gaps at the edges of the cow's body, as depicted in Figure 6. Compared to the original input depth image, these areas of depth gaps are relatively small in size. Therefore, in this stage, the Joint Bilateral Filtering (JBF) algorithm is employed to rectify the depth gaps within the depth image, thereby enhancing its overall quality. This step serves to mitigate measurement errors resulting from missing depth values at feature points.

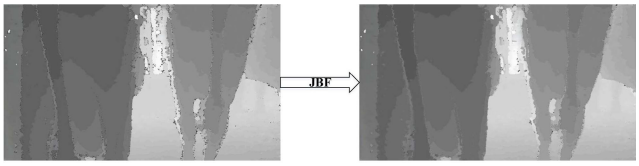


FIGURE 6. Data quality improvement.

2) BACKGROUND SEGMENTATION

The U-Net is a deep learning model specifically designed for image segmentation, characterized by its encoding-decoding structure, which allows it effectively utilize information from the input image by compressing it into smaller feature maps. Furthermore, the bottom-up architecture of U-Net enables it to identify details of larger objects. In a previous investigation, we optimized the U-Net neural network model using the PyTorch framework (version 1.5.0) to achieve non-contact background segmentation of livestock in the side-view. The optimized U-Net neural network is depicted in Figure 7.

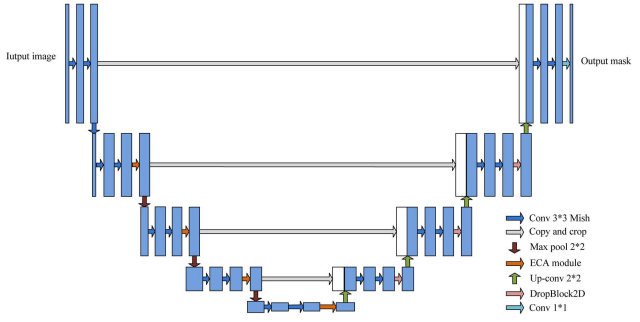


FIGURE 7. Segmentation model base on U-Net.

Initially, ReLU was used as the activation function for the U-Net neural network. However, due to the inherent limitation of ReLU, where the negative semi-axis value is always zero, the neurons can become inactive when the input of the model is at its minimum value. To overcome this issue, the ReLU activation function after each convolutional layer and batch normalization layer in the U-Net neural network was replaced with the Mish activation function. The Mish activation function avoids saturation due to function capping and the vanishing gradient and dead ReLU problem, as the positive axis derivative of Mish is greater than one. Moreover, the Mish algorithm has better generalization than the ReLU algorithm because it is smoother. Additionally, the DropBlock2D [32] module was incorporated into the up sampling process of U-Net. The DropBlock module is a regularization module for convolutional neural networks that effectively removes certain semantic information by randomly blocking out a portion of continuous regions. By acting as an effective regularize, the DropBlock module forces the network to learn other features of the object after several iterations of training. Furthermore, previous research suggests that appropriate cross-channel interaction could reduce the model complexity while maintaining performance [33]. Hence, after each down-sampling convolution module of U-Net, an attention module was added to enhance the accuracy of the segmentation.

However, there are still some shortcomings in this method. Specifically, the model was exclusively trained on cow data captured from a side view, which may compromise its performance when applied to images captured from different angles. Furthermore, due to lack of necessary image augmentation techniques, the ability of the model to segment the background at varying distances and angles is also limited. To address these limitations and improve the segmentation performance of the model, we employed the following strategies, as depicted in Figure 8. Firstly, we increased the number and variety of images in the training dataset, including images captured from different angles and distances. Secondly, we implemented a range of image augmentation techniques, such as sharpening, affine transformation, random pixel removal, and Gaussian noise addition during the image pre-processing stage, to enhance the ability of the model to extract salient features from images and improve its generalization capacity.

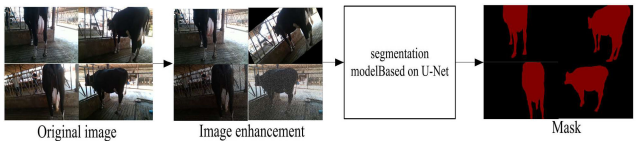


FIGURE 8. Background segmentation process.

After many experiments, a set of optimal parameter settings was determined to achieve more accurate cow segmentation. First, the Batch size was set to 20. Secondly, this study implemented a dynamic adjustment strategy, in which the initial learning rate was set to 0.01, and then reduced to 0.001 during the first 60-90 rounds to refine the model training. After 90 rounds, the learning rate was further reduced by a factor of 10 to prevent oscillation and slow convergence. Finally, the Stochastic Gradient Descent method was used as

the optimization algorithm, and the early stopping strategy was employed to prevent over fitting.

3) FEATURE POINT LOCATION

Enhancing the precision of feature point localization holds substantial importance in augmenting the accuracy of body condition scoring. However, there is a multitude of obstacles and intricacies, such as intricate scenes, modifications in scale, and diverse poses. The excellent solution for feature point location is to convert the regression problem to pixel-level classification problem based on heat map response. Referring to this idea, the present research has employed HRNet for feature point localization.

HRNet is a 2D human pose estimation network architecture, which was proposed by Sun [34] of CSU and Microsoft Asia Research in 2019. The current mainstream approach for multi-scale feature extraction typically involves down sampling high-resolution feature maps to low resolution, then up sampling the low-resolution feature maps back to high resolution [35], [36], [37]. However, traditional feature extraction approaches that extract features from high to low image resolutions lead to a loss of valuable information [23], [38], [39]. To address this defect, HRNet employs parallel sub-networks that maintain high-resolution feature image representations throughout the feature extraction process. At the end of each stage, the network fuses the feature maps of different resolutions. The feature maps from all resolutions are merged through a fusion layer to create the final feature representation. Each body joint feature point has a corresponding output channel, which produces a heat map for that specific joint feature point. This research introduced the Convolutional Block Attention Module (CBAM) [40] to elevate the weight of significant features in the channel and spatial axes while suppressing unnecessary features during the extraction process. Figure 9 depicts the HRNet structure based on the CBAM.

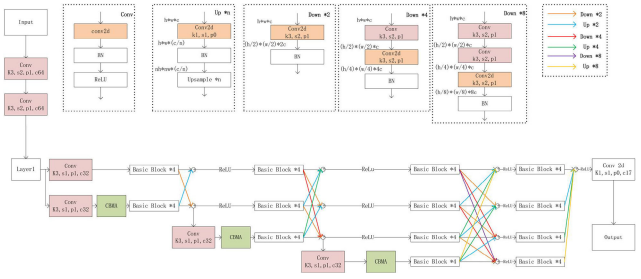


FIGURE 9. The structure of CBMA- HRNet.

HRNet is composed of four stages. In the first stage, the resolution is reduced to 1/4 of the input image through two convolutional layers with a 3*3 convolutional kernel, and the stride is set to 2. The regression heat map is represented at this resolution. The high-resolution subnet, consisting of repeatedly stacked Bottlenecks, is responsible for adjusting the number of image channels. Subsequently, the second, third, and fourth stages add subnets that comprise a series of transition structures and stage structures. Each transition structure introduces a new scale branch, doubling the number of channels while decreasing the resolution. The number of channels increases exponentially as the number of subnets in the network increases. Simply adding channel information directly ignores the correlation between channels. In addition, a CBMA module was add after each channel number expansion to enhance the weight of important features during the extraction process. Different resolution subnets were parallel connected, and repeated information fusion was performing at various scales. Ultimately, the position of the feature point was predicted through the high-resolution heat map.

CBAM is an attention mechanism module that integrates both spatial and channel attention. It comprises the Channel Attention Module (CAM) and the Spatial Attention Module (SAM). The SAM can be utilized within HRNet to effectively highlight significant spatial features within an image. HRNet is distinguished by its utilization of multiple resolution feature maps, where low-resolution feature maps capture global information, while high-resolution feature maps preserve local details. By strategically applying spatial attention across the various resolution feature maps in HRNet, it becomes possible to accentuate positions within these maps, consequently enhancing the precision of feature point localization. Similarly, CAM can be employed in HRNet to emphasize the significance of different channels. HRNet incorporates multiple branches, with each branch responsible for extracting features at distinct scales. Through the application of channel attention on the diverse branch feature maps in HRNet, it becomes feasible to highlight crucial channel features across different branches, thereby bolstering the robustness of feature point localization. The structure of CAM is shown in Figure 10.

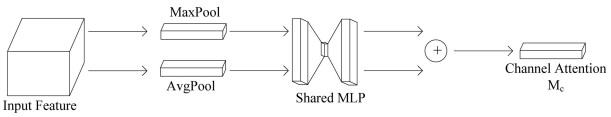


FIGURE 10. Structure of CAM.

Initially, the feature map with dimensions H*W*C was computed utilizing the MaxPool and AvgPool method, leading to a feature with dimensions of 1*1*C. Subsequently, the shared MLP with the ReLU activation function processed the two feature maps. After the element-wise addition of the two features, the resulting feature was subjected to a sigmoid activation function to obtain the channel attention feature $M_c(F)$. The calculation formula for $M_c(F)$ is expressed in equation (1), where $W_0 \in R^{c/(r*c)}$ and $W_1 \in R^{c*(\frac{c}{r})}$ are the weight of the hidden layer and output layer, correspondingly.

$$\begin{aligned} &M_c(F) \\ &= sigmoid\left(MLP\left(AvgPool\left(F\right)\right) + MLP\left(MaxPool\left(F\right)\right)\right) \\ &= sigmoid(W_1\left(W_0\left(F^c_{avg}\right)\right) + W_1\left(W_0\left(F^c_{max}\right)\right)) \end{aligned} \tag{1}$$

The structure of SAM is shown in Figure 11. For the input feature map, the SAM performs max pooling and average pooling operations along the channel dimension, respectively, to obtain a feature map with contextual information of different scales. Subsequently, the processed result F^s_{avg} and F^s_{max} is stacked on the same dimension, and then a 1*1 convolution is used to adjust the number of channels to generate spatial attention weights. Finally, use the sigmoid function to scale to between 0 to 1 and obtain the final spatial attention feature M_s. The formula for this feature is presented in equation (2). The convolution kernel was set to 7*7.

$$M_s(F) = sigmoid\left(f^{7*7}([AvgPool(F); MaxPool(F)])\right) = sigmoid(f^{7*7}([F^s_{avg}; F^s_{max}])) \quad (2)$$

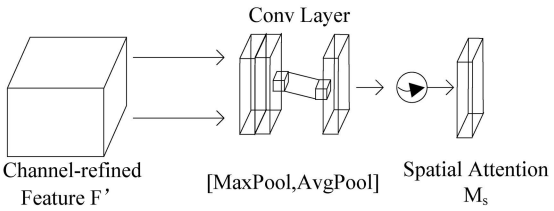


FIGURE 11. Structure of SAM.

F. EXPERIMENTAL SETTING

This study conducted two experiments to assess the efficacy of the proposed method and investigate the influence of distance, depth image quality, and a various of depth image restoration method on measurement precision.

Initially, the research aimed to evaluate the effect of distance on measurement accuracy, for which ten cow videos were randomly selected. From these videos, three color images and depth images were chosen for each feature point, based on suitable postures and complete depth values at the distances of 1-2 m, 2-3 m, and 3-4 m. The proposed method was utilized to score each feature and the average machine measurement was taken as the final feature score. Finally, the difference between the machine measurement and the manual measurement was computed to determine the measurement error.

The second experiment aimed to address the issue of missing depth values in depth maps captured by depth cameras due to factors such as distance, highlights, or shadows. As the distance increases, the area of missing depth values also increases, making such data unsuitable for use. To address this limitation and evaluate the impact of depth image quality and different depth image restoration methods on measurement accuracy, the research selected three suitable color images and corresponding depth images for 10 cows for each feature at distances of 1m-2m and 2m-3m. The depth values at feature points were processed using three methods outlined in Section II-E1. Each feature was then evaluated using the method proposed in this research, and the machine's average score was taken as the final result of the scoring feature at the specific distance. The error between the machine score and the manual score was then calculated to determine the effectiveness of the depth image restoration methods.

III. RESULT AND DOSCUSSION

A. COMPARISON OF BACKGROUND SEGMENTATION PERFORMANCE OF COW

To validate the effectiveness of the proposed method for cow segmentation, this study selected five common performance metrics in foreground segmentation: Accuracy, Sensitivity, Specificity, Params and execution time. We compared our model with four classic foreground segmentation models: SegNet, UNet, DeeplabV3 and UNet++. The results are shown in Table 2.

TABLE 2. Comparison of background segmentation performance.

	Accuracy	Sensitivity	Specificity	Params(M)	Time(ms)
SegNet	0.85	0.80	0.87	0.79	17.64
UNet	0.87	0.84	0.89	0.83	23.53
DeeplabV3	0.90	0.89	0.91	0.87	18.22
UNet++	0.93	0.92	0.93	0.91	16.88
our	0.95	0.96	0.95	0.94	24.11

Our model achieved the highest accuracy of 0.95, substantially outperforming the second best model UNet++ at 0.93. Among all comparing models, our model exhibited the most significant performance gain, fully demonstrating its superiority in overall segmentation correctness. For sensitivity, our model also attained the highest score of 0.96, exceeding the second best DeeplabV3 by a large margin of 7 percentage points. This remarkable improvement shows our model's stronger capability in identifying foreground objects. Sensitivity reflects directly the model's foreground recognition capability, so this metric's prominent enhancement further verifies the effectiveness of our method. In terms of specificity, our model is on par with UNet++ at 0.95. This suggests that while improving foreground recognition, our model also maintained strong distinction for background regions. By integrating the CBMA module, our model has slightly more parameters than other models. Considering the significant segmentation improvement, this increase is acceptable. Our model is also slightly slower than other models, mainly due to the extra modules designed to enhance performance. This trade-off is reasonable. Hardware acceleration can be explored in the future to reduce time consumption. In summary, compared with other segmentation models, our model achieved noticeable improvement on all key metrics, with more accurate foreground detection and substantial background interference reduction. This fully demonstrated the validity and superiority of our proposed model.

B. COMPARISON OF KEYPOINT LOCALIZATION PERFORMANCE OF COW

To validate the effectiveness of the proposed method for cow keypoint localization, this study selected five common performance metrics: Precision, Recall, mAP, Params and execution time. We compared our model with three classic

keypoint localization models: Hourglass, CPN and Hrnet. The results are shown in Table 3.

TABLE 3. Comparison of feature point localization performance.

	Precision	Recall	mAP	Params(M)	Time(ms)
Hourglass	0.78	0.74	0.77	25.10	72.15
CPN	0.85	0.84	0.87	27.89	101.43
Hrnet	0.86	0.89	0.89	28.54	52.12
our	0.94	0.93	0.95	29.73	63.45

Our model achieved a high precision of 0.94, outperforming the second and third best models by large margins of 9.5% and 8.5% respectively. This demonstrates substantial improvements in localization accuracy. For recall, our model also notably exceeded the second and third best models, indicating our model was able to cover and localize more true keypoints while avoiding misses. In terms of the overall metric mAP, our score of 0.95 also surpassed other models by a large extent. This fully proves that our model attained the most significant performance gains in comprehensive keypoint localization quality. Regarding model complexity, the number of parameters in our model is on par with others. This means we realized remarkable performance gains without increasing computational burden, which is noteworthy. Our model also demonstrated state-of-the-art time efficiency, further suggesting the effects come from genuine model optimization instead of efficiency compromise. In summary, quantitative analysis clearly shows that compared to existing advanced techniques, our proposed keypoint localization model achieves noticeable improvements across all core evaluation metrics. This fully demonstrates the validity of our method.

C. THE INFLUENCE OF DISTANCE ON MEASUREMENT ACCURACY

The regression models and coefficient of determination (R^2) values for manual and machine measurements is present in Table 4. At the camera distance of 1-2 m from the cow, the R^2 values for chest depth, hip height, ischial width, parallelism of hind, hind leg curvature, length of teat, and depth of teat were 0.9018, 0.9653, 0.9873, 0.9926, 0.9719, 0.9437, and 0.958, respectively. These R^2 values all exceeded 0.9, indicating a high level of consistency between the two measurement methods. At the camera distance of 2-3 m from the cow, the R^2 values for chest depth, hip height, ischial width, parallelism of hind, and hind leg curvature ranged from 0.8 to 0.95. However, the R^2 values for length of teat and depth of teat were both less than 0.2. It is indicating a lack of correlation between manual and machine measurements for these two variables at this distance. At a camera distance of 3-4 m from the cow, all the correlation coefficients for body measurements were less than 0.6, indicating a negligible correlation between manual and machine measurements at this distance.

The error box line plot of manual and machine measurements at corresponding distances from the cow is present in Figure 12. Results showed that the relative errors of measurement for chest depth, hip height, ischial width, and hind leg curvature ranged from −5.93% to 5.38% at a camera distance of 1-3 m from the cow. However, for parallelism of hind, the average relative error of measurement at 2-3 m reached −9.84%, and the maximum measurement error was −15.89%, which exceeded the acceptable range of normal measurement error. Conversely, the relative error ranges for length of teat and depth of teat were within the acceptable range at 1-2 m, with an average relative error of 3.92% and 4.9%, respectively. But, the average relative error of measurement for both lengths of teat and depth of teat was greater than 200% at 2-3 m and 3-4 m, indicating that measurements at these distances are not credible. The average relative error was too large compared with other body scales, so they are not show in the Figure 7. Furthermore, at a camera distance of 3-4 m from the cow, the coefficient of determination for manual and machine measurements of hip height and hind leg curvature were 0.5196 and 0.5029, respectively. The mean measurement errors for chest depth and parallelism of hind were 10.61% and 31.25%, respectively, with coefficient of determination values of 0.1366 and 0.4376, indicating that measurements at this distance are also unreliable.

Figure 13 presents the heat map of manual and machine scoring errors for various cow features, where the manual and machine scores were transformed from the measurement results against Table 1. At a camera distance of 1-2 m from the cow, the scoring accuracy for most features, except for hind leg curvature and length of teat, was above 80% with a scoring error of 1. Although the accuracy for hind leg curvature and length of teat was lower compared to other features, the error was limited to 2. The high correlation between manual and machine measurements for each feature, as indicated by the coefficient of determination of 0.9719 and 0.9437, could reduce the scoring error via linear fitting. Similarly, at 2-3 m camera distance, the overall error range was within 1-2 points except for length of teat and depth of teat. The high correlation between the manual and machine measurement results for each feature indicates that they could serve as a reference basis. However, as shown in Figure 8(c), the scoring error was significant and unevenly distributed at a camera distance of 3-4 m from the cow, making it difficult to adjust the scoring accuracy using other means. The scoring results at this distance were not suitable as a reference basis.

Based on the experimental findings, when positioned the camera at a distance of 1-2 meters from the cow, the proposed method exhibited accurate measurement of all seven features. It proved that this method is satisfied actual production requirements. Similarly, at a distance of 2-3 meters, the scoring accuracy of all features except hind leg curvature and length of teat was deemed satisfactory. Which demonstrated the effectiveness and feasibility of the proposed method. The study further revealed a decrease in scoring accuracy with increasing distance, given that the depth camera relies on structured light technology to measure depth by calculating the time for infrared laser light to reflect back from the camera to the object. Consequently, both manual and machine-based

TABLE 4. Linear Regression Analysis of body measurement.

	1m-2m		2m-3m		3m-4m	
	Fitting equation	R^2	Fitting equation	R^2	Fitting equation	R^2
chest depth	y = 0.9522x + 0.0826	0.9018	y = 0.9577x + 0.0671	0.8064	y = 0.5697x + 0.6705	0.1366
hip height	y = 1.0373x - 6.0062	0.9653	y = 0.8371x + 23.035	0.9136	y = 0.5936x + 57.879	0.5196
ischial width	y = 0.9298x + 1.0663	0.9873	y = 0.8827x + 1.931	0.9344	y = 0.6441x + 6.5628	0.5544
parallelism of hind	y = 0.9399x + 0.5526	0.9926	y = 0.9545x - 1.0193	0.9471	y = 1.2938x - 11.809	0.4376
hind leg curvature	y = 1.0063x + 0.7014	0.9719	y = 1.1241x - 16.14	0.9372	y = 0.7345x + 40.415	0.5029
length of teat	y = 1.167x - 0.7567	0.9437	y = 2.8356x + 1.4316	0.1166	y = 6.082x - 4.5446	0.1073
depth of teat	y = 1.0109x + 0.3602	0.9580	y = -0.5647x + 26.574	0.0298	y = 0.5343x + 32.874	0.0067

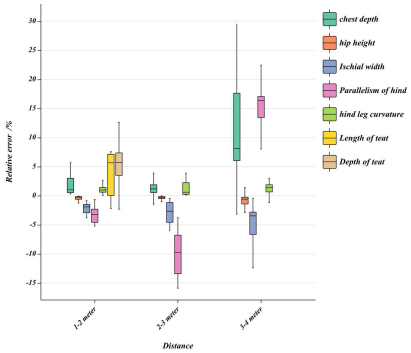


FIGURE 12. Error box line plot of manual and machine measurements at different distances.

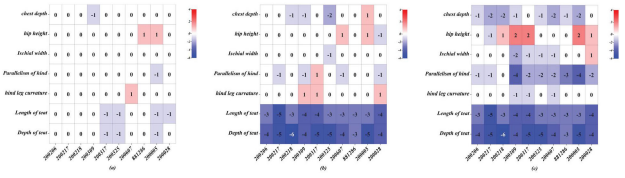


FIGURE 13. Heat map of manual scoring and machine scoring errors. (a), (b) and (c) are the error heat maps of 1-2 m, 2-3 m and 3-4 m camera distance from the cow, respectively.

measurements experienced a decline in accuracy in feature point localization and depth information acquisition, particularly when the cow's body size was reduced or the skeleton was unclear at larger distances. Specifically, measurements for hind leg curvature and length of teat required feature point P's localization and depth values on the color and depth images, respectively. Despite the localization accuracy of feature point P being satisfactorily determined through extensive annotation work.

The performance limitation of the depth camera led to weakened surface texture information at point P, thereby affecting the depth camera's ability to discern the depth of point P and the background, leading to distorted measurement results at distances ranging from 2-4 meters.

D. IMPACT OF DEPTH IMAGE RESTORATION METHODS ON MEASUREMENTS

Current depth image restoration methods can be broadly categorized into filtering, interpolation, function optimization, and deep learning approaches. Due to the absence of ground truth depth images of cows in practical production settings, supervised deep learning-based depth image completion methods are not applicable to this study. Therefore, this research opts to compare the proposed method with the Adaptive Autoregressive Model (AR) [41] based filtering, linear interpolation algorithm [10] and unsupervised GAN [42] to validate the reliability of the proposed approach.

The linear regression equations obtained from the manual and machine measurements of cow body features after utilizing depth image restoration method at a camera distance of 1-2 meters is presents in Table 5. As the feature of parallelism of hind and hind leg curvature were calculated by the special angle calculation of the line of sight points in the color image, their measurement results were solely dependent on the accuracy of feature point localization and not on the depth value of feature points. Therefore, these two features were excluded from the experiment. For the measurements of the remaining five cow body scales, namely, chest depth, hip height, ischial width, length of teat, and depth of teat, the correlation between manual and machine measurements exceeded 0.8 after applying the four restoration methods. This indicates that the missing depth regions near the feature point were small, and the three restoration methods provided consistent results in filling the small-scale depth deficiencies. A correlation of more than 0.75 signifies a high degree of interpretability and a good model fit. These results demonstrate the effectiveness of the proposed methods in achieving accurate measurements, even in the presence of depth deficiencies.

Table 6 presents the linear regression equations for machine and manual measurements after restoring the depth values of feature points of the depth image with three different methods for a camera distance of 2-3 m from the cow. In this research, parallelism of hind and hind leg curvature are not associated with depth values, so they were not calculated. Moreover, the length and depth of the teat features had low accuracy and high error rates in full depth image measurements when the camera was 2-3 m. As the camera distance increased to 2-3 meters, the overall fitting performance of the four methods in measuring key body dimensions of dairy cows decreased compared to the 1-meter distance condition, which is expected. Notably, the chest depth prediction of the

TABLE 5. Linear regression analysis of camera distance of 1-2 m.

	Lerp		AR		Gan		Our	
	Fitting equation	R^2	Fitting equation	R^2	Fitting equation	R^2	Fitting equation	R^2
chest depth	y = 1.0991x - 0.0889	0.8256	y = 1.0733x - 0.0724	0.8581	y = 1.461x - 0.5777	0.8643	y = 0.9111x + 0.1301	0.9102
hip height	y = 0.9992x - 1.3181	0.9515	y = 0.8496x + 20.705	0.8920	y = 0.7123x + 41.455	0.8533	y = 0.9389x + 8.313	0.9640
ischial width	y = 0.8773x + 1.8105	0.8414	y = 0.9056x + 1.4534	0.8824	y = 0.9623x + 0.8372	0.8533	y = 0.9148x + 1.2693	0.9314
length of teat	y = 0.9397x + 0.8416	0.9037	y = 0.914x + 0.8449	0.9204	y = 1.0369x - 0.1865	0.9048	y = 1.0378x - 0.0239	0.9173
depth of teat	y = 1.1524x - 0.8252	0.9407	y = 1.1132x - 0.473	0.9358	y = 0.8751x + 1.4203	0.9400	y = 1.0744x - 0.2214	0.9678

TABLE 6. Linear regression analysis of camera distance of 2-3 m.

	Lerp		AR		Gan		Our	
	Fitting equation	R^2	Fitting equation	R^2	Fitting equation	R^2	Fitting equation	R^2
chest depth	y = 0.5145x + 0.8954	0.2631	y = 0.9284x + 0.2658	0.4049	y = 0.982x + 0.0106	0.6844	y = 0.7496x + 0.3959	0.8263
hip height	y = 0.7201x + 33.972	0.5508	y = 1.0478x - 10.766	0.8028	y = 1.7363x - 108.04	0.7735	y = 1.19x - 29.989	0.9604
ischial width	y = 0.7115x + 4.3897	0.6991	y = 0.7278x + 4.0185	0.7413	y = 1.575x - 12.024	0.6415	y = 1.011x - 1.583	0.9244

Lerp method deteriorated the most drastically, with the R^2 dropping from 0.8256 to 0.2631, indicating its weak generalization capability. The results of the GAN method also showed some degree of performance decay at the 2-3 meter range, but our method demonstrated the strongest robustness. This implies the inadequacy of the three benchmark algorithms for regression tasks. Despite the adverse impact of lower image quality caused by the greater distance, our approach maintained chest depth, hip height and ischial width measurement accuracy with R^2 scores of 0.8263, 0.9604 and 0.9244 respectively, which were consistently higher than the other methods. Our framework could effectively learn descriptive features from the data, empowering the system with enhanced adaptability to the challenges of long-range imaging scenarios.

Figures 14(a) and 14(b) depict the accuracy of manual and machine measurements at two different camera distances, namely 1-2 m and 2-3 m, respectively. The depth images, which have undergone restoration using the Lerp model, the AR model, GAN and the method proposed in this study, exhibit relative measurement errors within the range of −10% to 10% for parameters including chest depth, hip height, Ischial width, Length of teat, and Depth of teat. The average relative errors of chest depth measurements using Lerp model, the AR model, GAN and the method proposed in this study were 2.63%, 1.40%, −1.14% and 1.75, respectively, and the average relative errors of hip height measurements were −0.98%, −0.81%, −0.28% and z0.40. Although the relative errors of length of teat and depth of teat measurements were larger compared to other features, considering the distortion of the measurements of these two features at other distances, and the correlation between manual and machine measurements of length of teat and depth of teat at 1-2 m using the four complementary methods was high. Therefore, the measurement results of cow body size features at 1-2 m using the depth images after the three complementary methods are acceptable.

The present research investigated the measurement errors of cow body features, including chest depth, hip height, and ischial width, using depth images processed by three different algorithms, at a camera distance of 2-3 m from the cow. The average relative errors of chest depth measurements using Lerp model, the AR model, GAN and the method proposed in this study were 24.68%, 14.58%, −10.93% and 7.33%, respectively, while the relative errors of hip height measurements were −4.65%, −2.61%, −1.81, and −1.61%. The mean relative errors for the ischial width measurements were −8.92%, −8.98%, 7.90%, and −6.09. It was observed that the measurement error increased with the distance. The results indicated that the Lerp model and the AR model were not suitable for measuring features when the camera is 2-3 m away from the cow. Although the GAN and the method proposed in this study showed a good fit for measuring chest depth and ischial width, the relative errors ranged from ±5% to ±10%, which was still not as accurate as the measurement results obtained at a camera distance of 1-2 m. Finally, for the measurement of hip height at 2-3 m, although there is a slight difference in measurement accuracy and fit compared to 1-2 m, the use of depth images processed by AR model, GAN and the method proposed in this study can still meet actual production requirements.

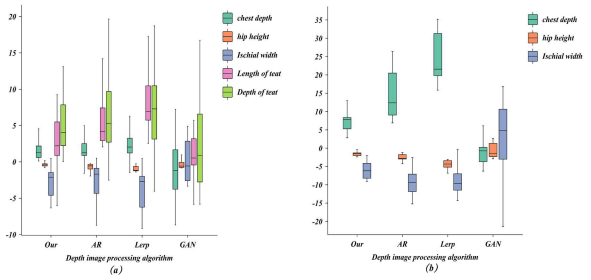


FIGURE 14. Error box line diagram for manual and machine measurements. (a) Is results for camera distance of 1-2 meters from the cow; (b) Is the result of the camera being 2-3 meters away from the cows.

As the distance increases between the camera and the cows, the areas of depth gaps in the depth images also expand. From the results above, it is evident that the AR algorithm, being based on filtering techniques, can only provide simple depth restoration for peripheral edges when faced with extensive depth gaps. The Lerp algorithm exhibits significant errors when dealing with functions with substantial curvature

or slope variations. Furthermore, Due to the limitations of the GAN network structure, it is difficult to capture high-frequency information, resulting in blurry generated outputs and ineffective recovery of detailed textures. Additionally, it has a notable impact on the smoothness of the depth image when non-uniform interpolation is employed. Therefore, constrained by the depth range within the depth image, when there are significant differences in depth, there is limited capacity for repair in the central vicinity of the missing depth areas.

Hence, as the distance increases, compared to the AR, Lerp and GAN model, the algorithm proposed in this study proves to be effective in handling extensive depth value gaps. The majority of feature points in this study are positioned near the edges of the cow's body. However, the depth image reconstruction algorithm proposed in this study establishes a framework for reverse sampling from a simple distribution, implements smooth transitions with the concept of time steps, and progressively controls noise removal. This method not only leverages the guidance of prior information from RGB images but also combines its own generative capabilities to achieve unsupervised high-quality depth image restoration.

Therefore, whether at distances of 1-2 meters or 2-3 meters, the measurement results obtained using the algorithm proposed in this study are superior to those of the other algorithms.

IV. CONCLUSION AND FUTURE STUDIES

This research presents a novel method for measuring body condition score for cow, which involves measuring and scoring seven features at various distances. The methodology of this research consists of three primary components.

Firstly, 19 feature points were localized using the HR-Net and the positioning locating correction method, coupled with an attention mechanism, based on the concept of human feature point localization; Secondly, employed a self-built database and image augmentation technique to segment background at different angles and distances; Thirdly, the measurement and scoring of seven cow features were completed. Additionally, we investigated the effects of distance, depth image quality, and different depth image restoration method on measurement and scoring accuracy.

The research evaluated the accuracy of a proposed method for measuring seven features of cows using depth images with complete depth values. Results indicated that chest depth, hip height, ischial width, parallelism of hind, hind leg curvature, length of teat, and depth of teat measurements achieved a coefficient of determination above 0.9 and relative measurement errors ranging from −5.93% to 8.33% when the camera distance of 1-2 m from the cow. After conversion into scores, the accuracy of feature scores exceeded 80%. These findings demonstrate the effectiveness and feasibility of the proposed method for meeting actual production requirements at this camera distance. For camera distances of 2-3 m, the coefficient of determination was greater than 0.8, and the average relative measurement error ranged between −9.84% and 1.26%, except for the length and depth of the teat. Overall, these results suggest that the proposed method is reliable at this distance and can apply to practical production.

In this study, the AR algorithm, Lerp algorithm, GAN and depth image completion method proposed in this study were evaluated for their ability to complete and measure depth values of cow feature points in the presence of missing data. The results showed that all four algorithms could effectively handle small-scale depth deficits for features such as chest depth, hip height, ischial width, length of teat, and depth of teat. Which a high degree of agreement between machine and manual measurements (coefficient of determination >0.8, and relative measurement error controlled within −10% to 10%) at a camera distance of 1-2 m from the cow. However, when the camera was 2-3 m away from the cow, the AR algorithm, GAN and Lerp algorithm algorithms showed larger differences in measurement accuracy than the method proposed in this study, which was able to effectively handle large-scale depth missing. Therefore, the method proposed in this study is the most suitable for completing measurements at both 1-2 m and 2-3 m distances, as it can meet the actual production requirements and ensure accurate results.

The present research successfully utilized deep learning methods to measure seven features. However, given that there are 20 scoring items for cows, there is room for future improvement by expanding the measurement and scoring scope to include additional items. Additionally, future studies should explore more advanced techniques for segmenting and localizing the hoof and udder of cows. Currently, the measurement and scoring process only utilizes image processing techniques, but future research will explore video streaming techniques to improve accuracy and completeness.

REFERENCES

[1] W. R. Butler and R. D. Smith, "Interrelationships between energy balance and postpartum reproductive function in dairy cattle," *J. Dairy Sci.*, vol. 72, no. 3, pp. 767–783, Mar. 1989.

[2] G. Azzaro, M. Caccamo, J. D. Ferguson, S. Battiato, G. M. Farinella, G. C. Guarnera, G. Puglisi, R. Petriglieri, and G. Licitra, "Objective estimation of body condition score by modeling cow body shape from digital images," *J. Dairy Sci.*, vol. 94, no. 4, pp. 2126–2137, Apr. 2011.

[3] A. Bercovich, Y. Edan, V. Alchanatis, U. Moallem, Y. Parmet, H. Honig, E. Maltz, A. Antler, and I. Halachmi, "Development of an automatic cow body condition scoring using body shape signature and Fourier descriptors," *J. Dairy Sci.*, vol. 96, no. 12, pp. 8047–8059, Dec. 2013.

[4] I. Halachmi, M. Klopcic, P. Polak, D. J. Roberts, and J. M. Bewley, "Automatic assessment of dairy cattle body condition score using thermal imaging," *Comput. Electron. Agricult.*, vol. 99, pp. 35–40, Nov. 2013.

[5] J. Salau, J. H. Haas, W. Junge, U. Bauer, J. Harms, and S. Bieletzki, "Feasibility of automated body trait determination using the SR4K time-of-flight camera in cow barns," *SpringerPlus*, vol. 3, no. 1, pp. 1–16, Dec. 2014.

[6] J. M. Bewley, A. M. Peacock, O. Lewis, R. E. Boyce, D. J. Roberts, M. P. Coffey, S. J. Kenyon, and M. M. Schutz, "Potential for estimation of body condition scores in dairy cattle from digital images," *J. Dairy Sci.*, vol. 91, no. 9, pp. 3439–3453, Sep. 2008.

[7] I. Halachmi, P. Polak, D. J. Roberts, and M. Klopcic, "Cow body shape and automation of condition scoring," *J. Dairy Sci.*, vol. 91, no. 11, pp. 4444–4451, Nov. 2008.

[8] S. Battiato, G. M. Farinella, G. C. Guarnera, G. Puglisi, G. Azzaro, and M. Caccamo, "Assessment of cow's body condition score through statistical shape analysis and regression machines," in *Proc. 1st Workshop Appl. Pattern Anal.*, 2010, pp. 66–73.

[9] X. Huang, Z. Hu, X. Wang, X. Yang, J. Zhang, and D. Shi, "An improved single shot multibox detector method applied in body condition score for dairy cows," *Animals*, vol. 9, no. 7, p. 470, Jul. 2019.
[10] R. Spoliansky, Y. Edan, Y. Parmet, and I. Halachmi, "Development of automatic body condition scoring using a low-cost 3-dimensional Kinect camera," *J. Dairy Sci.*, vol. 99, no. 9, pp. 7714–7725, Sep. 2016.
[11] S. Yukun, H. Pengju, W. Yujie, C. Ziqi, L. Yang, D. Baisheng, L. Runze, and Z. Yonggen, "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.*, vol. 102, no. 11, pp. 10140–10151, Nov. 2019.
[12] A. Fischer, T. Luginbühl, L. Delattre, J. M. Delouard, and P. Faverdin, "Rear shape in 3 dimensions summarized by principal component analysis is a good predictor of body condition score in Holstein dairy cows," *J. Dairy Sci.*, vol. 98, no. 7, pp. 4465–4476, Jul. 2015.
[13] J. Rodríguez Alvarez, M. Arroqui, P. Mangudo, J. Toloza, D. Jatip, J. M. Rodríguez, A. Teyseyre, C. Sanz, A. Zunino, C. Machado, and C. Mateos, "Body condition estimation on cows from depth images using convolutional neural networks," *Comput. Electron. Agricult.*, vol. 155, pp. 12–22, Dec. 2018.
[14] D. Liu, D. He, and T. Norton, "Automatic estimation of dairy cattle body condition score from depth image using ensemble model," *Biosystems Eng.*, vol. 194, pp. 16–27, Jun. 2020.
[15] B. M. Martins, A. L. C. Mendes, L. F. Silva, T. R. Moreira, J. H. C. Costa, P. P. Rotta, M. L. Chizzotti, and M. I. Marcondes, "Estimating body weight, body condition score, and type traits in dairy cows using three dimensional cameras and manual body measurements," *Livestock Sci.*, vol. 236, Jun. 2020, Art. no. 104054.
[16] M. F. Hansen, M. L. Smith, L. N. Smith, K. A. Jabbar, and D. Forbes, "Automated monitoring of dairy cow body condition, mobility and weight using a single 3D video capture device," *Comput. Ind.*, vol. 98, pp. 14–22, Jun. 2018.
[17] J. D. Ferguson, D. T. Galligan, and N. Thomsen, "Principal descriptors of body condition score in Holstein cows," *J. Dairy Sci.*, vol. 77, no. 9, pp. 2695–2703, Sep. 1994.
[18] D. Anglart, "Automatic estimation of body weight and body condition score in dairy cows using 3D imaging technique," Dept. Animal Nutrition Manag., Faculty Vet. Med. Animal Sci., Swedish Univ. Agricult. Sci., Tech. Rep., 2010, doi: 10.13140/RG.2.2.26909.26084.
[19] Y. Kuzuhara, K. Kawamura, R. Yoshitoshi, T. Tamaki, S. Sugai, M. Ikegami, Y. Kurokawa, T. Obitsu, M. Okita, T. Sugino, and T. Yasuda, "A preliminarily study for predicting body weight and milk properties in lactating Holstein cows using a three-dimensional camera system," *Comput. Electron. Agricult.*, vol. 111, pp. 186–193, Feb. 2015.
[20] K. Zhao, A. N. Shelley, D. L. Lau, K. A. Dolecheck, and J. M. Bewley, "Automatic body condition scoring system for dairy cows based on depth-image analysis," *Int. J. Agricult. Biol. Eng.*, vol. 13, no. 4, pp. 45–54, 2020.
[21] Y. Le Cozler, C. Allain, A. Caillot, J. M. Delouard, L. Delattre, T. Luginbuhl, and P. Faverdin, "High-precision scanning system for complete 3D cow body shape imaging and analysis of morphological traits," *Comput. Electron. Agricult.*, vol. 157, pp. 447–453, Feb. 2019.
[22] T. T. Zin, P. T. Seint, P. Tin, Y. Horii, and I. Kobayashi, "Body condition score estimation based on regression analysis using a 3D camera," *Sensors*, vol. 20, no. 13, p. 3705, Jul. 2020.
[23] W.-Y. Li, Y. Shen, D.-J. Wang, Z.-K. Yang, and X.-T. Yang, "Automatic dairy cow body condition scoring using depth images and 3D surface fitting," in *Proc. IEEE Int. Conf. Unmanned Syst. Artif. Intell. (ICUSAI)*, Nov. 2019, pp. 155–159.
[24] A. Pezzuolo, M. Guarino, L. Sartori, and F. Marinello, "A feasibility study on the use of a structured light depth-camera for three-dimensional body measurements of dairy cows in free-stall barns," *Sensors*, vol. 18, no. 3, p. 673, Feb. 2018.
[25] A. N. Ruchay, K. A. Dorofeev, V. V. Kalschikov, V. I. Kolpakov, and K. M. Dzhulamanov, "A depth camera-based system for automatic measurement of live cattle body parameters," *IOP Conf. Ser., Earth Environ. Sci.*, vol. 341, no. 1, Oct. 2019, Art. no. 012148.
[26] A. Ruchay, V. Kober, K. Dorofeev, V. Kolpakov, and S. Miroshnikov, "Accurate body measurement of live cattle using three depth cameras and non-rigid 3-D shape recovery," *Comput. Electron. Agricult.*, vol. 179, Dec. 2020, Art. no. 105821.
[27] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
[28] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, P.-A. Manzagol, and L. Bottou, "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion," *J. Mach. Learn. Res.*, vol. 11, no. 12, pp. 3371–3408, 2010.
[29] K. Jiang, Z. Wang, P. Yi, G. Wang, T. Lu, and J. Jiang, "Edge-enhanced GAN for remote sensing image superresolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 8, pp. 5799–5812, Aug. 2019.
[30] Y. Xiao, Q. Yuan, K. Jiang, J. He, X. Jin, and L. Zhang, "EDiffSR: An efficient diffusion probabilistic model for remote sensing image super-resolution," 2023, *arXiv:2310.19288*.
[31] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 6840–6851.
[32] G. Ghiasi, T.-Y. Lin, and Q. V. Le, "Dropblock: A regularization method for convolutional networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 31, 2018, pp. 10750–10760.
[33] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, "ECA-Net: Efficient channel attention for deep convolutional neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 11531–11539.
[34] K. Sun, B. Xiao, D. Liu, and J. Wang, "Deep high-resolution representation learning for human pose estimation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 5686–5696.
[35] K. Jiang, Z. Wang, P. Yi, C. Chen, G. Wang, Z. Han, J. Jiang, and Z. Xiong, "Multi-scale hybrid fusion network for single image deraining," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 7, pp. 3594–3608, Jul. 2023, doi: 10.1109/TNNLS.2021.3112235.
[36] Y. Xiao, X. Su, Q. Yuan, D. Liu, H. Shen, and L. Zhang, "Satellite video super-resolution via multiscale deformable convolution alignment and temporal grouping projection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–19, 2022, Art. no. 5610819, doi: 10.1109/TGRS.2021.3107352.
[37] K. Jiang, Z. Wang, P. Yi, C. Chen, B. Huang, Y. Luo, J. Ma, and J. Jiang, "Multi-scale progressive fusion network for single image deraining," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 8343–8352.
[38] Z. Tian, H. Chen, and C. Shen, "DirectPose: Direct end-to-end multi-person pose estimation," 2019, *arXiv:1911.07451*.
[39] X. Nie, J. Feng, J. Zhang, and S. Yan, "Single-stage multi-person pose machines," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 6950–6959.
[40] S. Woo, J. Park, J. Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 3–19.
[41] J. Yang, X. Ye, K. Li, C. Hou, and Y. Wang, "Color-guided depth recovery from RGB-D data using an adaptive autoregressive model," *IEEE Trans. Image Process.*, vol. 23, no. 8, pp. 3443–3458, Aug. 2014.
[42] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proc. Neural Inf. Process. Syst.*, 2014, pp. 1–9.

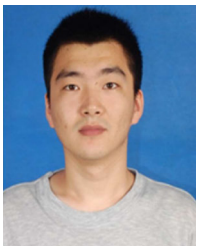
KEQIANG LI received the B.S. degree in software engineering from the Shenyang University of Technology, Shenyang, China, in 2014, and the M.S. degree in software engineering from Yanshan University, Qinhuangdao, China, in 2017. He is currently pursuing the Ph.D. degree in mechanical and electrical engineering with Hebei Agricultural University, Baoding, China. He joined the School of Mathematics and Information Technology, Hebei Normal University of Science and Technology, Qinhuangdao, in 2014. His research interests include computer vision and deep learning.

GUIFA TENG is currently the Dean of the College of Information Science and Technology, Hebei Agricultural University. He is also a Professor and a Doctoral Supervisor.

JIANTAO WANG received the master's degree in agriculture from Hebei Agricultural University, in 2008, majoring in animal genetic breeding and reproduction. He has been with Tangshan Animal Husbandry Technology Promotion Station, since 2012. He is currently a Senior Livestock Breeder. He is mainly responsible for the scientific research and promotion of new technology, new breeds, and new achievements in livestock and poultry.

YUXIN ZHANG received the bachelor's degree majoring in animal medicine from Jilin University, in December 2006. He is currently a Senior Veterinarian. He has been with Tangshan Animal Husbandry Technology Promotion Station, since 2016, mainly responsible for the technical guidance of livestock and poultry breeding and scientific research and promotion of new technologies, new breeds and new achievements, technology, new varieties, and new achievements in scientific research and promotion.

LEI GAO received the master's degree in agronomy from Jilin Agricultural University, in 2011, majoring in preventive veterinary medicine. He is currently a Senior Veterinarian. He is mainly responsible for livestock structural adjustment, standardized production, large-scale rearing, and livestock statistics. His main research interests include the infectious diseases of livestock and poultry and their prevention and control.



HUI FENG received the bachelor's degree in animal science from Jiangnan University, in 2010, and the B.S. degree in agricultural science, in 2022. He has been with Tangshan Animal Husbandry Technology Promotion Station, mainly responsible for the promotion of breeding technology of breeding livestock and poultry. His research interests include livestock and poultry breeding and breeder management.