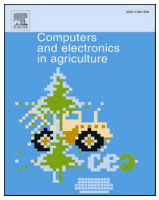
Computers and Electronics in Agriculture 229 (2025) 109773

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

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



Original papers

# Cow depth image restoration method based on RGB guided network with modulation branch in the cowshed environment

Yanxing Li [a], Xin Dai [a], Baisheng Dai [a,*], Peng Song [a], Xinjie Wang [a], Xinchao Chen [a], Yang Li [b], Weizheng Shen [a,*]

[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

ARTICLE INFO

*Keywords:*
Dairy cow
Depth image restoration
RGB guidance
Deep learning
Body condition score

ABSTRACT

Depth images were widely applied in smart animal husbandry. The raw depth images collected by the RGB-D cameras generally existed amount of missing depth values due to the light reflected from white pattern of cows and direct sunlight in the cowshed. The incomplete cows in depth images would affect the application of depth images in health monitoring. This study proposed a cow depth image restoration method based on RGB guided network with a modulation branch. Firstly, removing the outliers caused by light from the depth image and determining the depth value missing area of the cow's body. Second, RGB and depth features were extracted through multiple convolutions and fused in the S-C (Self-attention and Convolution attention) fusion module of encoder. Then, the prediction head generated a coarsely predicted depth image after deconvolution combined with a modulation branch. Finally, the repaired depth image was generated in the SPN (Spatial Propagation Network) refinement module of the decoder. In terms of dataset construction, 7260 depth images were collected in a commercial dairy farm. To make up for lacking ground truth complete depth images corresponded to the raw depth images with missing value, two ways for generating missing depth images were designed. The experimental results shown that the method had improved restoration quality of cow's incomplete body in depth images. By comparing with other depth restoration works, the proposed method achieved significantly superior performance on RMSE = 36.32 and MAE = 12.77, and the percentage of predicted pixels within the error range at 1.25 reached 0.999. Additionally, a smoother transition between missing and restoration regions was demonstrated in the repaired depth images and point cloud results. And compared with the depth images with missing regions, the Precision, Recall rate and F1-score of the repaired depth images were improved for cow body condition scoring. This study could improve the effectiveness of the collected data and make the depth images more practical for smart animal husbandry.

## 1. Introduction

The depth image data had been widely applied in the field of smart animal husbandry (He et al., 2023; Tan et al., 2023) and correct pixel-wise object depth value played a substantial role in cow body measurement, weight estimation, and body condition scoring (Li et al., 2022; Shen et al., 2023; Shi et al., 2023; Yang et al., 2022). However, during the process of generating a depth image, the camera was often affected by the sunlight and the movement of the object in the pasture environment, causing the collected depth image contained large amounts of missing pixels. The lack of depth values resulted in the loss of depth information, which affected subsequent applications. Therefore, it was necessary to study the depth image restoration.

Existing depth image restoration methods were categorized into traditional depth image restoration method and deep learning-based depth image restoration method. The traditional depth image restoration method mainly used filters to fill holes and artifacts in depth images. In order to avoid the distortions and artifacts caused by hole for cow body condition scoring, Sun et al. had took the weighted joint bilateral filter (Sun et al., 2019). The filter was designed to average the pixels that were spatially near and have similar intensity values by a jointly computed kernel on a RGB image with a weight map (Matsuo et al., 2013; Petschnigg et al., 2004). Several studies had used filters in spatial and temporal dimension to fill holes. In order to achieve cow

* Corresponding authors.
*E-mail address:* bsdai@neau.edu.cn (B. Dai).

https://doi.org/10.1016/j.compag.2024.109773
Received 5 August 2024; Received in revised form 23 November 2024; Accepted 2 December 2024
Available online 12 December 2024

body condition scoring based on RGB and depth data, Winkler et al. firstly took spatial edge preserving filter to fill holes, and then a temporal filter was applied to frame sequence of one cow to reduce the differences between frames (Winkler et al., 2024). Jia et al. adopt two steps to repair and fill the cow's depth value missing caused by random noise. In the first step, the mode-filter was used to process depth images in spatial dimension and the second step was using a weighted averaging filter in temporary dimension (Jia et al., 2021). Shen et al. found that the missing of depth values had an impact on cow's weight estimation. Using hole filling filter in spatial dimension and temporal filter could complete the holes and the accuracy of cow's weight estimation had been improved (Shen et al., 2023). In addition to the methods mentioned above, there were also interpolation (Atapour-Abarghouei and Breckon, 2017; Garro et al., 2009) and other image processing techniques (Islam et al., 2017; Lu et al., 2014; Park et al., 2014) applied to hole filling in depth images (Xie et al., 2024). However, methods described above lacked sufficient capability to fill the missing regions with complex shape based on valid pixels.

Deep learning-based restoration method could better utilize multi-scale or contextual information of depth images to repair the missing areas with irregular shape. Therefore, a considerable amount of literature had been studied on depth image restoration based on deep learning. The depth image restoration methods can be mainly classified into unguided depth image restoration and RGB guided depth image restoration (Hu et al., 2023). The unguided depth image restoration methods contained sparsity-aware CNN (Chodosh et al., 2019; Uhrig et al., 2017), normalized CNN (Eldesokey et al., 2020; Eldesokey et al., 2018) and training with auxiliary images (Lu et al., 2020; Yu et al., 2021), which aimed at directly completing depth images using deep neural network. But the unguided methods suffered from blur effect and distortion of object boundaries due to the lack of prior information such as neighboring objects and sharp edges.

The RGB images could provide information about object structures, including textures, lines and edges, thus, a growing number of works (Dimitrievski et al., 2018; Jaritz et al., 2018; Ma and Karaman, 2018; Ryu et al., 2021) employed RGB guided method to repair depth images. Zhang et al. concatenated the RGB and depth features in the channel dimension, and then connected the convolutional attention and Transformer in parallel to fuse the features. This fusion method was beneficial to the feature interaction between local and global context information (Zhang et al., 2023). Huang et al. combined RGB image with depth image in channel dimension as the input, and adopt the self-attention mechanism to enhance the feature meanings of spatial location and channel map. Then the gated convolution (Yu et al., 2019) was used to predict more accurate depth values (Huang et al., 2019). Gansbeke et al. designed a two branches framework, including global branch and local branch. Specifically, the RGB and depth image concatenated in channel dimension was used to generated guidance map and global depth prediction in global branch, and the depth image fused guidance map in the local branch to generate local depth prediction. At the end of the framework, added the two predicted results to obtain the final depth image (Gansbeke et al., 2019). Through feature fusion, the RGB information was introduced to repair the irregularly shaped missing regions by providing more prior information. To better predict the depth image, Wang et al. adopted a two-branch end-to-end network, and this study propagated depth information to RGB image by assigning the same weight value to regions with the same depth value, and assign the same weight to regions with the same color to smooth the depth image (Wang et al., 2022). The contour and color information provided by RGB images ensured that irregularly shaped missing areas had clear boundaries and smoother depth values. Although, methods described above could repair missing areas with irregular shape in depth image, incorrect depth value predictions occurred in the cowshed environment. The reason was that the prior information provided by RGB image was insufficient due to light reflection of cow's white pattern.

Therefore, to enhance the utilization effect of prior information from RGB image and improve the accuracy of depth value prediction, this study proposed a cow depth image restoration method based on RGB guided network with modulation branch. Firstly, the outliers caused by sunlight was removed to reduce the impact on depth image quality. Then, the features of RGB and depth images were fused through the S-C (Self-attention and Convolution attention) fusion module in the encoder. Next, the SPatially-Adaptive DEnormalization (SPADE) block was used to establish the relationship between the fused features and missing depth values through mapping features to the modulation signal generated by input mask of missing depth values. After a series of deconvolution, the coarsely predicted depth image was generated in the prediction head. Finally, the depth image was further refined after spatial propagation in the SPN (Spatial Propagation Network) refinement module to generate the final depth image in the decoder. Overall, the contributions were summarized as follows: (1) Innovatively introducing a method for cow depth image restoration in the cowshed environment, which contained outlier removal and a RGB guided network with modulation branch. (2) The SPADE block was used to establish feature mapping relationship between the fused features and missing depth values to improve the utilization effect of RGB image. (3) A cow depth image restoration dataset had been publicly released.

## 2. Materials and method

### 2.1. Data collection

Experimental multi-view data of Holstein dairy cows were collected in a commercial farm named Sheng Kang farm, which located in Da Qing City, Heilongjiang Province, China. Data of 121 Holstein dairy cows were captured from 13 July to 5 August in 2022. Multi-view data collection system was placed on passageway of the experimental cowshed as shown in Fig. 1. The collection system consisted of metal framework, five Intel Real-sense D435 cameras and two laptop computers. The metal framework was designed according to the previously work of Li et al. (Li et al., 2022). The Intel Real-sense D435 camera offered quality depth for a variety of applications and can be integrated into any solution with ease. The ideal range of Intel Real-sense D435 camera was 0.3 m to 3 m, depth output resolution was up to 1280 × 720 pixels and depth frame rate was up to 90 fps. The two laptop computers hardware configuration of this collection system were Intel ( R ) Core ( TM ) i7-10875H@2.30 GHz with 16 GB RAM and 12th Gen Intel(R) Core(TM) i5-12500H@2.50 GHz with 16 GB RAM. The installation position of cameras was shown in Fig. 1 (a,b,c,d,e) and multi-view data were simultaneously recorded on the laptop computer via USB 3.0 cables.

In order to collect the data of Holstein dairy cows, a multi-view data collection algorithm was designed. Specifically, the collection algorithm was based on 64-bit Windows 11 system, Visual Studio 2019 and Intel RealSense SDK 2.47. To ensure that cameras collected data simultaneously, the single view collection algorithm provided by Intel RealSense SDK was modified. When a cow passed through the multi-view data collection system, five cameras started to capture cow data and saved it as bag file at same times. As shown in Fig. 2, the raw depth images and RGB images were extracted from bag files through Python 3.7. The cow depth image restoration dataset was publicly released at https://github.com/IPCLab-NEAU/Cow-Depth-Image-Restoration. It can be seen that depth value missing generally existed in the side and back views of raw depth images. The reasons that result in depth value missing could be roughly categorized into three types:

(1) Direct sunlight, sun shined directly on the D435 camera, causing partial depth value of the cow side was missing as shown in Fig. 2 (a);
(2) Intense sunlight reflection, depth value missing existed in depth images of the Holstein cow because the strong sunlight reflection as shown in Fig. 2(b) and (c);


**Fig. 1.** The multi-view data collection system.

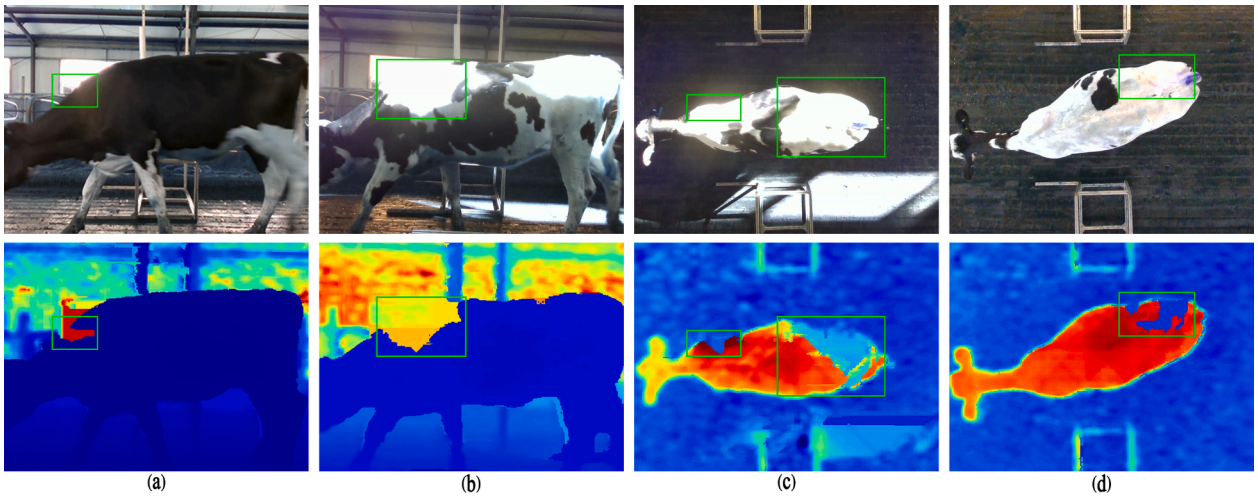



**Fig. 2.** RGB images and depth images extracted from bag file. (a) Depth value missing caused by direct sunlight, (b) depth missing of intense sunlight reflection in side view, (c) depth missing of intense sunlight reflection in back view, (d) depth missing of sunlight reflection of white pattern, (depth images had converted to pseudo color images and green box indicated the position of missing regions).

(3) Sunlight reflection, the white patterns on the back of cow also caused depth value missing under the normal lighting condition in the cowshed as shown in Fig. 2(d).

The lacking of depth values could affect the subsequent application of depth images, it was necessary to repair the missing regions.

### 2.2. Methodology pipeline

The main workflow of this study was summarized in Fig. 3. To complete body missing region of cows, a cow depth image restoration method including outlier removal, determination of cow boundaries based on Mask2Former and RGB guided depth image restoration

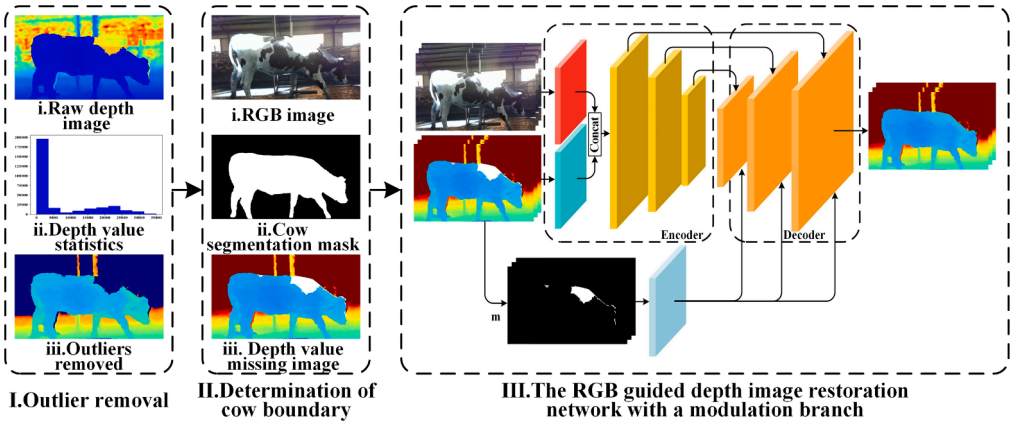



**Fig. 3.** The workflow of RGB guided depth image restoration method.

network was proposed. The outlier removal was designed to map the outliers to a reasonable value using depth value statistics. The determination of cow boundaries was used to set the depth values to zero in the missing areas of cow's body. The RGB guided depth image restoration network fused the features of RGB and depth images through attention mechanism in encoder and achieved depth image restoration through a modulation branch in the decoder. Details of this workflow were described as follows.

### 2.3. Outlier removal

Due to direct sunlight in the side view, there were some highlight areas existed in the window areas of the RGB image and depth value of the highlight areas was an outlier as shown in Fig. 4(a) and (b). These outliers would affect the feature representation of cow body data in depth images, resulting in insignificant differentiation between depth values on the surface of cow body. To reduce the impact of outliers, firstly, the range of outliers was determined using depth histograms of depth images. Then, the depth threshold was determined based on the range. At last, the outliers were set to a reasonable value based on the equation (1), the result was shown in Fig. 4(c).

$$\begin{cases} d_{(i,j)\in(H,W)} = d_{(i,j)\in(H,W)}, if\, d_{(i,j)\in(H,W)} \leqslant d_{\text{thresh}} \\ d_{(i,j)\in(H,W)} = d_{\text{thresh}}, \text{others} \end{cases} \tag{1}$$

where $d_{(i,j)}$ was the depth value in $(i, j)$, $H$ and $W$ were the height and width of depth image, $d_{\text{thresh}}$ was the depth threshold for removing outliers, the $d_{\text{thresh}}$ was set as 4500 to ensure that the depth value range of depth images from different views was the same in this study.

### 2.4. Determination of cow boundaries based on Mask2Former

In the depth images, the missing depth values caused by lighting were non-zero outliers. It was necessary to set these outliers as zero to determine the cow's body boundary. Mask2former demonstrated superiority in instance segmentation involving cows. Therefore, a Mask2-Former network was trained using pre-trained weights for instance segmentation of cows in this study to obtain cow boundaries. Firstly, the Mask2Former was trained based on RGB images to segment cows from the background. The depth values within the cow segmentation mask were set to zero based on equation (2) in the depth images as shown in Fig. 5.

$$\begin{cases} d_{(i,j)\in(H,W)} = 0, if\, d_{(i,j)\in(H,W)} = d_{\text{thresh}} \text{ and } (i,j) \in (m,n) \\ d_{(i,j)\in(H,W)} = d_{(i,j)\in(H,W)}, \text{others} \end{cases} \tag{2}$$

where $(m, n)$ were the vertical and horizontal coordinates of the cow segmentation mask.

### 2.5. RGB guided depth image restoration network with a modulation branch

The structure of proposed depth image restoration network was shown in Fig. 6. The RGB and depth images were encoded in the convolution layers and two ResNet34 layers. The S-C fusion modules were used to fuse the features and accomplished the final encoding. In

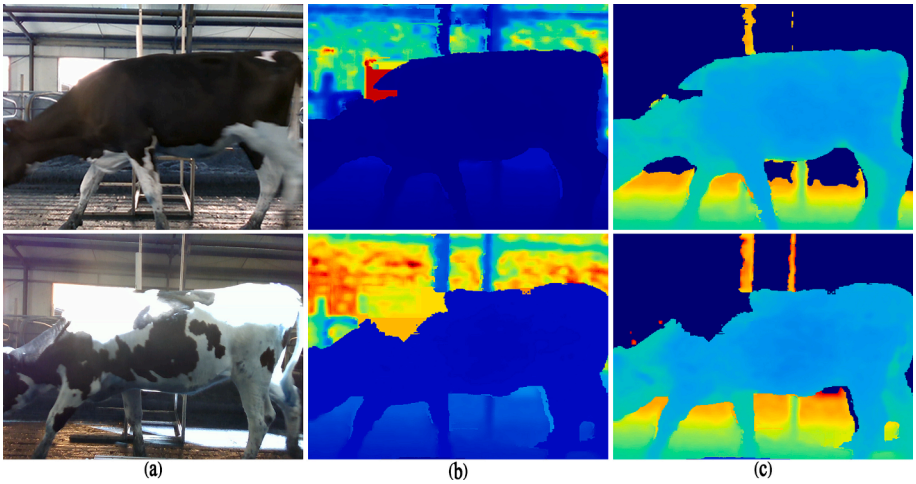



**Fig. 4.** Outlier depth value removal. (a) the RGB images, (b) raw depth images, (c) depth images after outlier removal.

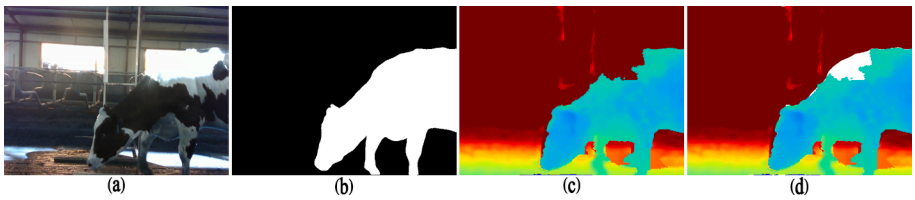



**Fig. 5.** Setting depth values of missing area to zero based on instance segmentation. (a) was the RGB images, (b) was the cow segmentation mask, (c) was the raw depth images, (d) was the depth images setting depth value to zero in missing areas.

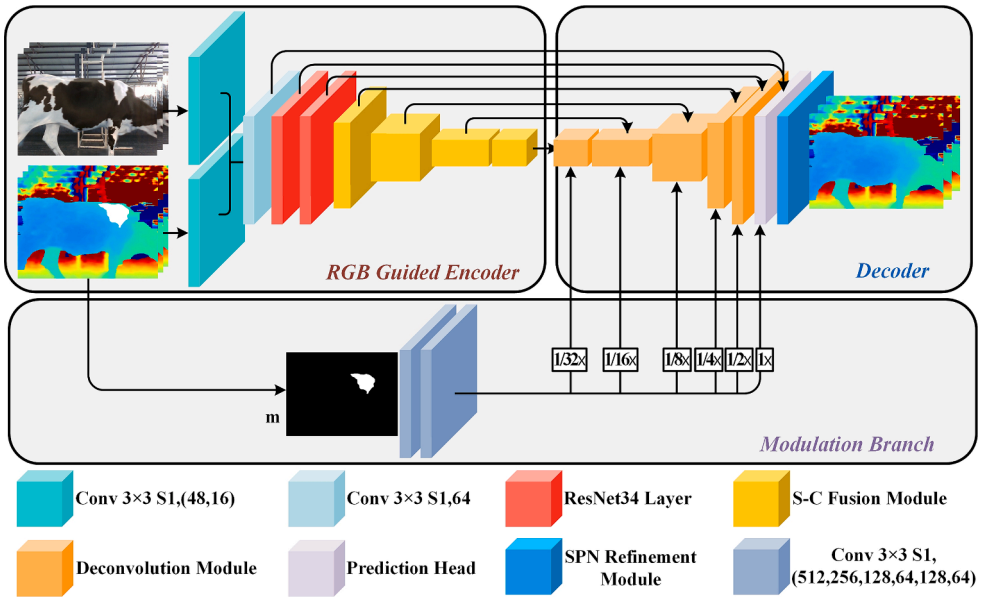



**Fig. 6.** Structure of RGB guided depth image restoration network with a spatially adaptive modulation branch.

the decoder, the features were deconvoluted combined with the modulation branch. At last, the feature maps were used to obtain the coarsely predicted depth image in the prediction head, and then refined multiple times in SPN refinement module to generate final depth image.

### 2.5.1. The RGB guided encoder

The RGB guided encoder, as shown in Fig. 7, the encoder with three stages to fuse the features. Four S-C fusion modules were added after the ResNet34 layers to enhance the ability to extract modality-specific information, which included Self-attention module and Convolution attention module. Detailed network architecture was shown in Fig. 8. The input features $F_i^{\text{in}} \in \mathbb{R}^{C_i \times H_i \times W_i}$ was delivered to the S-C fusion module to generate the output $F_i^{\text{out}} \in \mathbb{R}^{2C_i \times \frac{H_i}{2} \times \frac{W_i}{2}}$, where $i \in (1,2,3,4)$ was the $i$ th layer of S-C fusion module, the $C$, $H$, $W$ were the channel, height and width of the features.

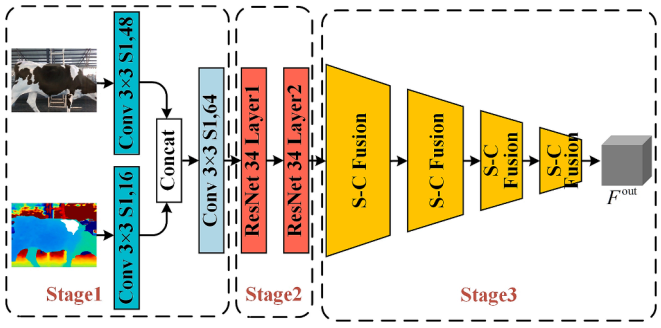



**Fig. 7.** The RGB guided encoder architecture.

### 2.5.2. The decoder with a modulation branch

The decoder included deconvolution modules combined with a SPatially-Adaptive DEnormalization (SPADE) block (Senushkin et al., 2021), a prediction head and a SPN refinement module as shown in Fig. 9.

The SPADE block was added before deconvolution as shown in Fig. 10(a). The feature maps and modulation signal **m** generated by zero values from missing depth image were inputted into SPADE block. Fig. 10(b) and (c) shown the specific structure of SPADE. Let $f_{n,c,y,x}^i \in \mathbb{R}^{Ci \times Hi \times Wi}$ denote the activation map of the $i$ th layer of the decoder for a batch of $N$ samples. The output value from the SPADE at site ($n \in N$, $c \in C_i$, $y \in H_i$, $x \in W_i$) was shown in equation (3).

$$D_{n,c,y,x}^i = \gamma_{n,c,y,x}^i(\mathbf{m}) \frac{f_{n,c,y,x}^i - \mu_c^i}{\sigma_c^i} + \beta_{n,c,y,x}^i(\mathbf{m}) \tag{3}$$

where $\mu_c^i = \frac{1}{N_i H_i W_i} \sum_{n,y,x} f_{n,c,y,x}^i$ was the mean and $\sigma_c^i = \sqrt{\frac{1}{N_i H_i W_i} \sum_{n,y,x} \left(f_{n,c,y,x}^i - \mu_c^i\right)^2}$ was the standard deviation, $\gamma_{n,c,y,x}^i$ and $\beta_{n,c,y,x}^i$ were the learned spatially dependent scale and bias parameters of the batch normalization layer. The SPADE mapped the fused features of RGB and depth images to the modulation signal generated by input mask of missing depth values, ensuring that the depth values generated in missing areas were uniform and smooth like RGB image.

The coarsely predicted depth image was generated after two convolutions in the prediction head as shown in Fig. 10(d). At last, the refined depth image was generated after $k$ steps spatial propagation in the SPN refinement module.

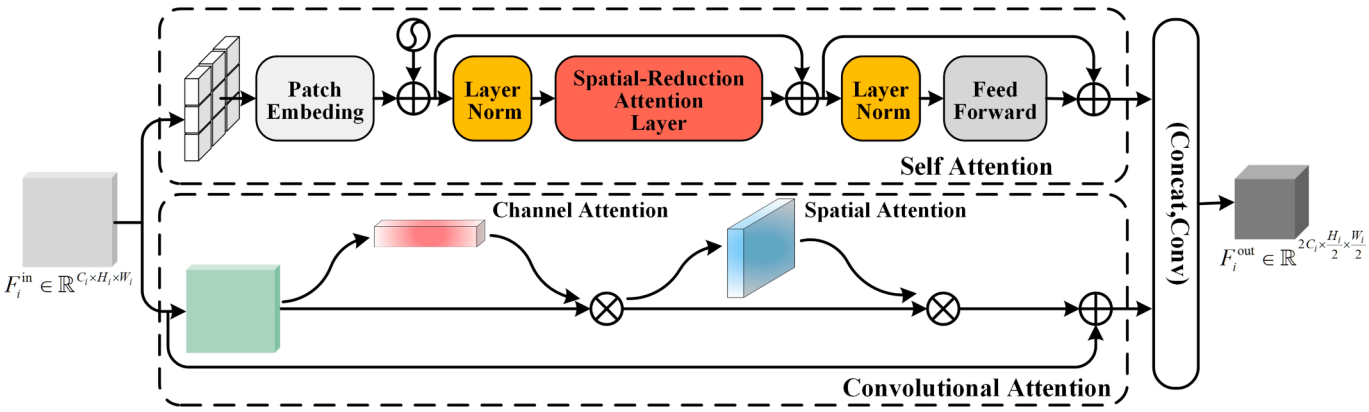



**Fig. 8.** The RGB and depth image fusion module based on self-attention and convolution attention.

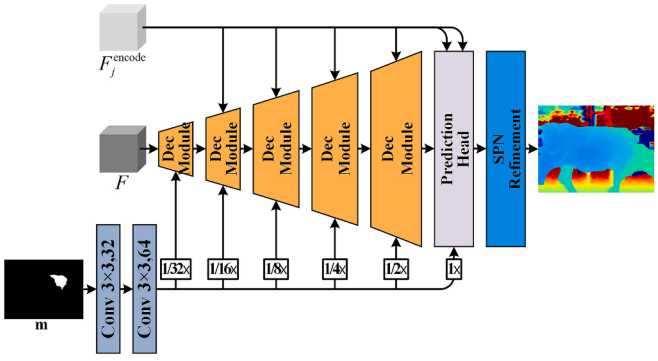



**Fig. 9.** The decoder with a deconvolution module combined with a modulation branch.

## 3. Experiments

### 3.1. Dataset construction

#### 3.1.1. Generate depth images with missing areas

During the construction of dataset, there were no ground truth complete depth images corresponded to the depth images with missing regions. In order to train the restoration network, there were two ways to generate depth images with missing areas based on the complete depth image, including Mask R-CNN model and channel value of images.

Inspired by the previous work of Wang et al. (Wang et al., 2022), a missing area generate algorithm was using for producing missing region on the complete depth images. As shown in Fig. 11, the first and second columns simulated the depth value missing caused by direct sunlight and sunlight reflection, while the third and fourth columns simulated the depth value missing caused by sunlight reflection. The depth value of the generated missing area was zero, represented by the white area in the Fig. 11(c). Considering the restoration for scene with missing area in the depth image was meaningless for production, this study only focused on cow's body region. The depth images with missing areas were generated as depth camera data and the complete depth images were used as ground truth data for the training.

This study took the Mask R-CNN to generate missing depth values, the process was shown in Fig. 12.

Randomly select 20 % depth images and the corresponding RGB images from the complete cow images as the train set. Manually annotated the area of cows that were easily missing depth values in the RGB images as the segmentation mask utilizing semi-automatic image segmentation and annotation tool ISAT (Interactive Segmentation Annotation Tool). Then the Mask R-CNN semantic segmentation model was trained by the annotated RGB images. Inputting other RGB images into the trained model to generate segmentation masks. Some generated segmentation masks would cover a portion of the background. It was necessary to remove the segmentation mask that covered the background based on the boundaries of cow's body mentioned in section 2.4. At last, mapping the segmentation mask back to the depth image to generate missing regions. The segmentation results were shown in Fig. 11.

To further expand the depth missing dataset of cows, the other way based on channel values was proposed. Specifically, setting the pixels of depth image as zero which R, G, and B channel value of corresponding to RGB image equal 255 and the depth values of depth image was in the range of 1300 to 1330.

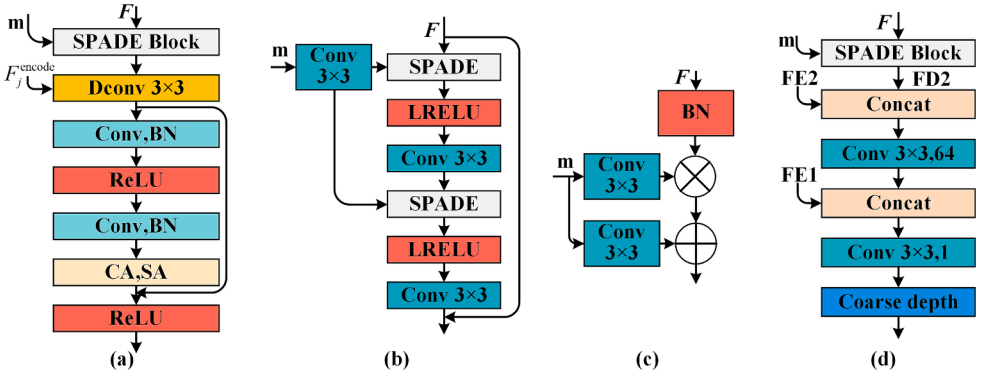



**Fig. 10.** The modules of decoder. (a) Deconvolution module, **m** was the modulation signal, ***F*** was the fused features from encoder and $F_j^{\text{encode}}$ was the feature from $j \in (1,2,3,4,5,6)$ th layer of encoder, (b) SPADE block, (c) SPADE, (d) Prediction head, **FE1** and **FE2** were the output from the second and third layer of encoder, the **FD2** was the output from the last layer of deconvolution module.

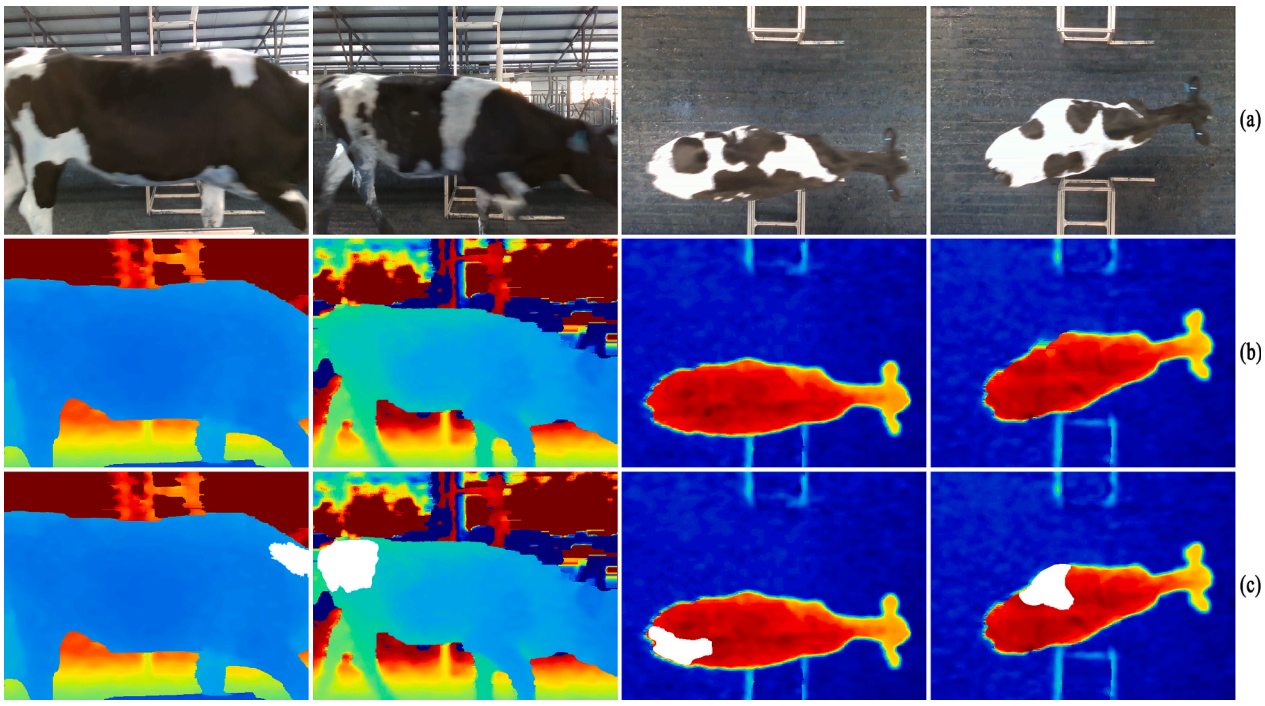



**Fig. 11.** The depth images with missing areas generated by Mask R-CNN. (a) was the RGB images, (b) was the complete depth images, (c) was the depth images with generating missing areas.


**Fig. 12.** Process of generating missing depth values.

#### 3.1.2. Dataset

In the experiment, the Manual Cow Depth Missing Dataset, Real Cow Depth Missing Dataset and BCS Test Dataset were constructed and all depth datasets mentioned below had corresponding RGB images. The Manual Cow Depth Missing Dataset constructed based on missing generation method described above, including data from cow's views of side and back. The Real Cow Depth Missing Dataset constructed by depth images with missing depth values collected in the cowshed. This dataset also contained data from cow's views of side and back. The BCS Test Dataset only contained data from the back view, as well as the body condition score of each cow. The partition of datasets was shown in Table 1.

### 3.2. Experiment setup

The experiment was conducted on the Ubuntu 18.0.4 operating system. The hardware configuration included an Intel(R) Xeon(R) CPU E5-2678 v3 @ 2.50 GHz, 128 GB of memory, a 1 TB solid-state drive, and 8 NVIDIA GeForce RTX 3090 GPUs, each with 24 GB of VRAM. The CUDA 11.4, PyTorch 1.10.0 and Python 3.8 operating environment were built on this basis. The original images were resized to 320 × 240 pixels as the input to the network, and the base learning rate, batch size, and epoch were set to 0.0001, 4, and 500, respectively.

**Table 1**
The cow dataset distributions.

| Dataset | Train set | | Test set | |
|---|---|---|---|---|
| | Side | Back | Side | Back |
| Manual Cow Depth Missing Dataset | 6776 | 3388 | 2904 | 1452 |
| Real Cow Depth Missing Dataset | / | / | 584 | 624 |
| BCS Test Dataset | / | 1456 | / | 364 |

### 3.3. Evaluation metrics

In order to assess method performance, three evaluation metrics were used as below. Given ground truth depth $D_0$ and restoration depth $D$, the metrics including (Huang et al., 2019):

(1) RMSE (Root Mean Square Error):

$$\sqrt{\frac{1}{|\text{cows}|}\sum_{\text{cow}\in\text{cows}}\|D(\text{cow})-D_0(\text{cow})\|^2} \tag{4}$$

(2) MAE (Mean Absolute Error):

$$\sqrt{\frac{1}{|\text{cows}|}\sum_{\text{cow}\in\text{cows}}\|D(\text{cow})-D_0(\text{cow})\|} \tag{5}$$

(3) (3) $\delta_t$: the percentage of predicted pixels within the error range $t$, the error range was defined as:

$$\max\left(\frac{D(\text{cow})}{D_0(\text{cow})},\frac{D_0(\text{cow})}{D(\text{cow})}\right)<t \tag{6}$$

where $t\in\{1.05,1.10,1.25\}$, cow $\in$ cows.

### 3.4. Comparisons with general restoration methods

Pre-experiments had shown that traditional methods cannot fill large missing areas. To verify the effectiveness of the proposed method, a comparison was made with deep learning-based methods. Since there was no depth restoration method based on deep learning applied to cows currently, we had fine-tuned three general restoration methods named

CompletionFormer (Zhang et al., 2023), NLSPN (Park et al., 2020) and DM-LRN (Senushkin et al., 2021). The restoration performance of the four methods was tested on the Manual Cow Depth Missing Dataset. The quantitative comparison results were shown in Table 2.

Compared to all other works, the proposed method had significantly superior performance on RMSE = 36.32, MAE = 12.77 and won the highest score in $\delta_t$ ($t = 1.05, 1.10, 1.25$). When the threshold $t = 1.25$, the $\delta_t$ of our method was 0.999. Compared to CompletionFormer, the proposed method had achieved a better result in all evaluation metrics, indicating the efficacy of the SPADE block. At the same time, it was verified that the proposed method had a better ability to utilize prior information compared to other works. The $\delta_t$ of NLSPN was only 0.001 lower than CompletionFormer at $t = 1.25$, and higher than CompletionFormer at $t = 1.05$ and 1.10. All evaluation metrics of DM-LRN were the lowest among the four methods, but there was still an ideal result of 0.994 at $t = 1.25$.

The visualization results were shown in Fig. 13, It was apparent from this figure that the proposed method received clearer and smoother results. Rows 2 and 4 represent magnified area from rows 1 and 3, respectively. From this visualization results, we can see that DM-LRN resulted in the worst restoration outcome for cow's side when the depth images were missing large depth values, and only simple depth values of targets were generated by the DM-LRN. From the regions of cow's body in Fig. 13, the depth value variation of proposed method between the restoration area and the complete area was smoother. Overall, NLSPN and DM-LRN could repair the missing areas, but it ignored the continuity and smoothness of the depth values of the cow's body.

### 3.5. Ablation studies and analysis

In this section, we demonstrated the effectiveness of each component proposed in this method, including the S-C fusion module, the SPADE block and SPN refinement module.

**The S-C fusion module**. As can be seen from the Table 3, the RGB guided encoder with S-C fusion module significantly achieved higher RMSE, MAE and $\delta_t$ than the encoder without S-C fusion module. $\delta_{1.10} = 0.977$ was the most improved evaluation metric in $\delta_t$. And in the visualization result, Fig. 14 shown the contour of the missing area was not obvious and the missing depth values of targets were more similar to those in the complete areas. The addition of S-C fusion module in the encoding stage improved the prediction accuracy of depth values.

**The SPADE block**. Table 3 presented the results of adding SPADE block. From the data above, we can see that the network had a significantly increased by 8.18 and 2.88 respectively in RMSE and MAE. As shown in Fig. 14, it was apparent that the area of the darker part in the pseudo color image had decreased, gradually matching the color of the complete area. This indicates that the generation of incorrect depth values had decreased. This benefited from the SPADE block and its enhanced depth features in the network.

**The SPN refinement module**. From the Table 3, it can be seen that after SPN refinement all the evaluation metrics were increased. Further analysis showed that it resulted in the greatest improvement in RMSE and $\delta_{1.05}$ compared with other evaluation metrics. The accuracy of the depth values generation acquired a further improvement after refinement. From the visual results, it shown that the restoration region become smoother than the network without the SPN refinement module.

**Table 2**
The quantitative comparison results with general restoration methods.

| | RMSE | MAE | $\delta_{1.05}$ | $\delta_{1.10}$ | $\delta_{1.25}$ |
|---|---|---|---|---|---|
| CompletionFormer | 41.83 | 15.09 | 0.954 | 0.978 | 0.997 |
| NLSPN | 46.28 | 17.27 | 0.977 | 0.983 | 0.996 |
| DM-LRN | 60.13 | 24.50 | 0.952 | 0.969 | 0.994 |
| Our | 36.32 | 12.77 | 0.981 | 0.989 | 0.999 |

This gained from the SPN refinement module utilized neighbor values with corresponding similarities to improve the smoothness of the transition between missing and complete region.

### 3.6. The visual results of conversion into point clouds

In order to demonstrate the effect of depth restoration more intuitively, the depth camera's internal parameters for each view were obtained through a bag file. The depth data was converted into point cloud data through the depth camera's internal parameter formula as follow:

$$\begin{cases} X = D_{\text{cow}} \times (u - u_0) \times \dfrac{dx}{f} \\ Y = D_{\text{cow}} \times (v - v_0) \times \dfrac{dy}{f} \\ Z = D_{\text{cow}} \end{cases} \tag{7}$$

where $X$ ,$Y$ and $Z$ were the point cloud coordinates, ($u$,$v$) were the depth image coordinates, ($u_0$,$v_0$) was coordinates of the intersection point between the camera's optical axis and the imaging plane, $D_{\text{cow}}$ was the depth value at ($u$,$v$), $dx$ and $dy$ represented the length of a single pixel in $x$ direction and $y$ direction, $f$ was the focal length.

The visualization results were shown in Fig. 15, rows 2 and 4 represent magnified area from rows 1 and 3. Further analysis of the visualization results revealed that the interior of the restoration area was continuous, with smooth transition between the missing and the complete area. There were no noises, overlaps and holes in the restoration areas. Rows 1 and 3 were the cow's back and neck point clouds with missing regions, comparing the two restoration results in rows 2 and 4, it can be seen that the restoration regions still maintain the original curvature.

### 3.7. Experiment on repairing depth images of cows in a cowshed environment

In order to verify the practicality of the proposed method, the proposed method was tested on Real Cow Depth Missing Dataset. Due to there were no pixel-level aligned complete depth image corresponding to missing depth image, this study only evaluated the visualized restoration results in Fig. 16.

From the results of the restoration depth image, it can be seen that the proposed method can effectively restore the body contour of cows. This benefited from the result of mapping the fused features to the mask generated by the missing depth values in the SPADE block. The SPN refinement module ensured that the depth values inside the missing area after restoration were continuous, and the transition between the restoration area and the complete area was smooth. From the results of the point cloud, it can be seen that the distribution of the restoration area point cloud was similar to that of the complete area. And the edge of restoration areas remained continuous and sharp. The transition between the restoration and complete areas was smooth, and there was no noise in the transition area. This indicated that the unevenness of outliers and transition areas had been greatly improved by the SPADE block and SPN refinement module.

### 3.8. Application experiment of cow depth restoration method

We mainly adopt the Precision, Recall rate and F1-score to evaluate the performance of body condition scoring model in raw and restoration depth images. First of all, repairing the raw depth images with holes and artifacts from BCS Test Dataset using the proposed method. Then, the restoration depth images were used to predict the body condition score of the cow based on EfficientNetV2, and compared with the results of the raw depth images.

As shown in Table 4 and Fig. 17, it can be seen that the Precision of

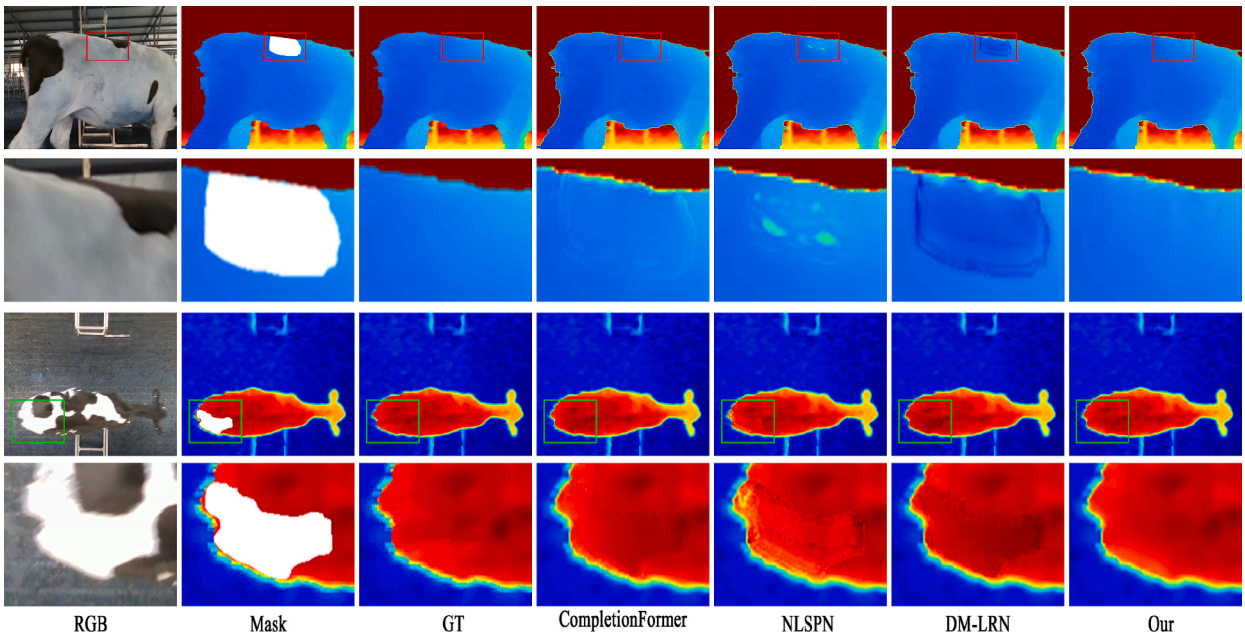



**Fig. 13.** Qualitative comparison with general restoration methods on test set.

**Table 3**
Ablation study on the Manual Cow Depth Missing Dataset.

| | RMSE | MAE | $\delta_{1.05}$ | $\delta_{1.10}$ | $\delta_{1.25}$ |
|---|---|---|---|---|---|
| RGB guided encoder | 79.57 | 39.09 | 0.955 | 0.961 | 0.985 |
| RGB guided encoder + S-C fusion | 48.63 | 20.78 | 0.962 | 0.977 | 0.993 |
| RGB guided encoder + S-C fusion + SPADE | 40.45 | 17.90 | 0.970 | 0.984 | 0.996 |
| RGB guided encoder + S-C fusion + SPADE + SPN module | 36.32 | 14.77 | 0.981 | 0.989 | 0.999 |

this model based on restoration depth images had improved compared to the raw depth images. Although the Recall rate had decreased in the body condition score of 3.75, for the body condition score of 2.5 and 4 were improved from zero to 14 % and 5 %. At the same time, the body condition score of 2.75 was still 5 %. For the F1-score in Table 4, all results had been improved compared to the raw depth images. The restoration results were shown in Fig. 18.

## 4. Conclusions

This paper proposed a cow depth image restoration method based on RGB guided network with modulation branch in cowshed environment. The proposed method could directly model relationship between fused features and missing area of depth image to reduce the prediction of incorrect depth values based on SPADE block. And the depth values of missing regions could be refined by SPN refinement module to improve the smoothness of the transition between restoration and complete regions. In addition, to better train the restoration network, a set of methods based on Mask R-CNN network and channel values of images were designed to generate depth images with missing depth value based on the complete depth images. The experimental results demonstrated the proposed method achieved a superior performance on RMSE = 36.32 and MAE = 12.77, the result of real missing depth images indicated a higher smoothness and sharpness. The results of cow body condition scoring based on restoration depth images shown that the proposed method can achieve improvements in Precision 9 %, Recall rate 9 % and F1-score 8 % within 0 error range. The future work would focus on restoring the cow's body surface with more realistic

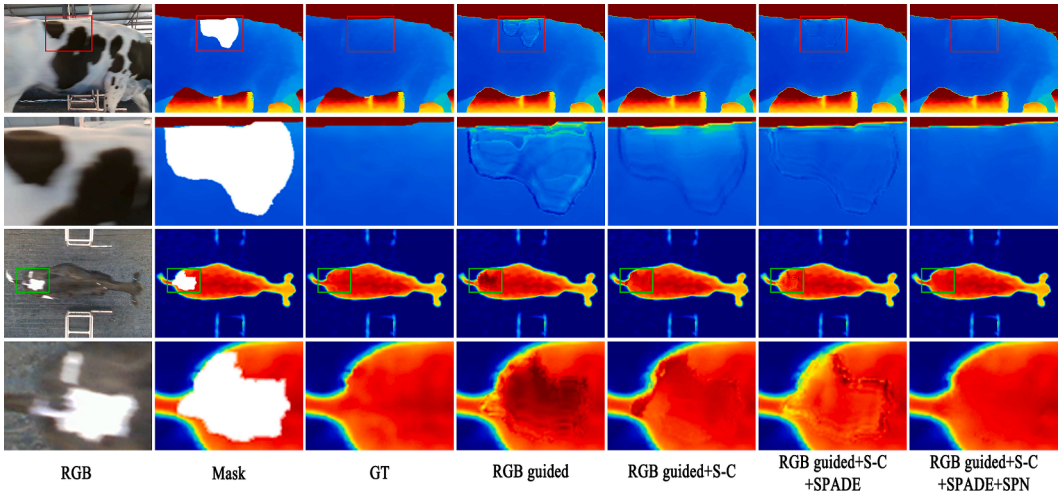



**Fig. 14.** The visualization results of ablation study on the Manual Cow Depth Missing Dataset.

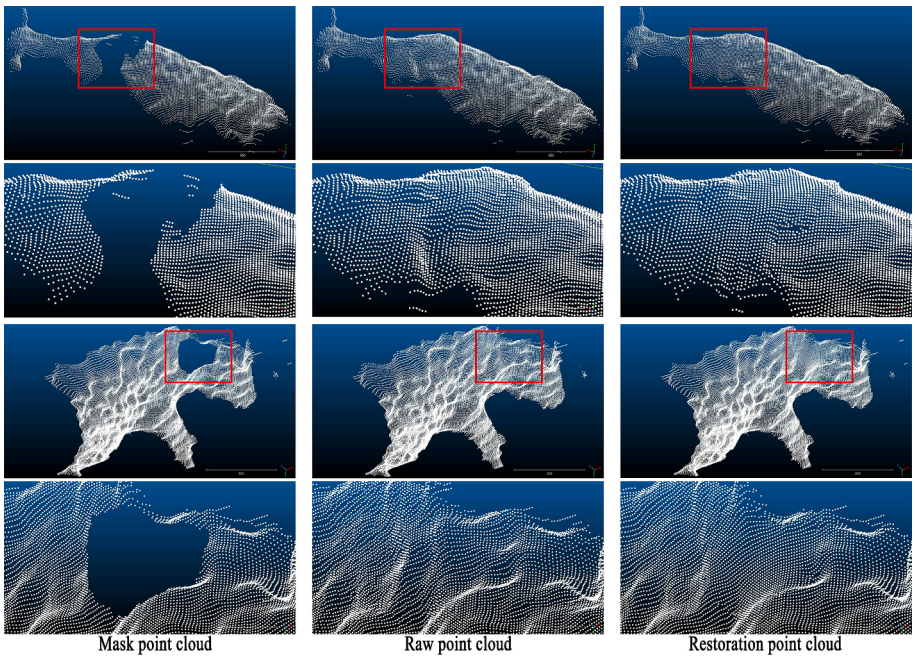



**Fig. 15.** The visualization results of conversion into point clouds.

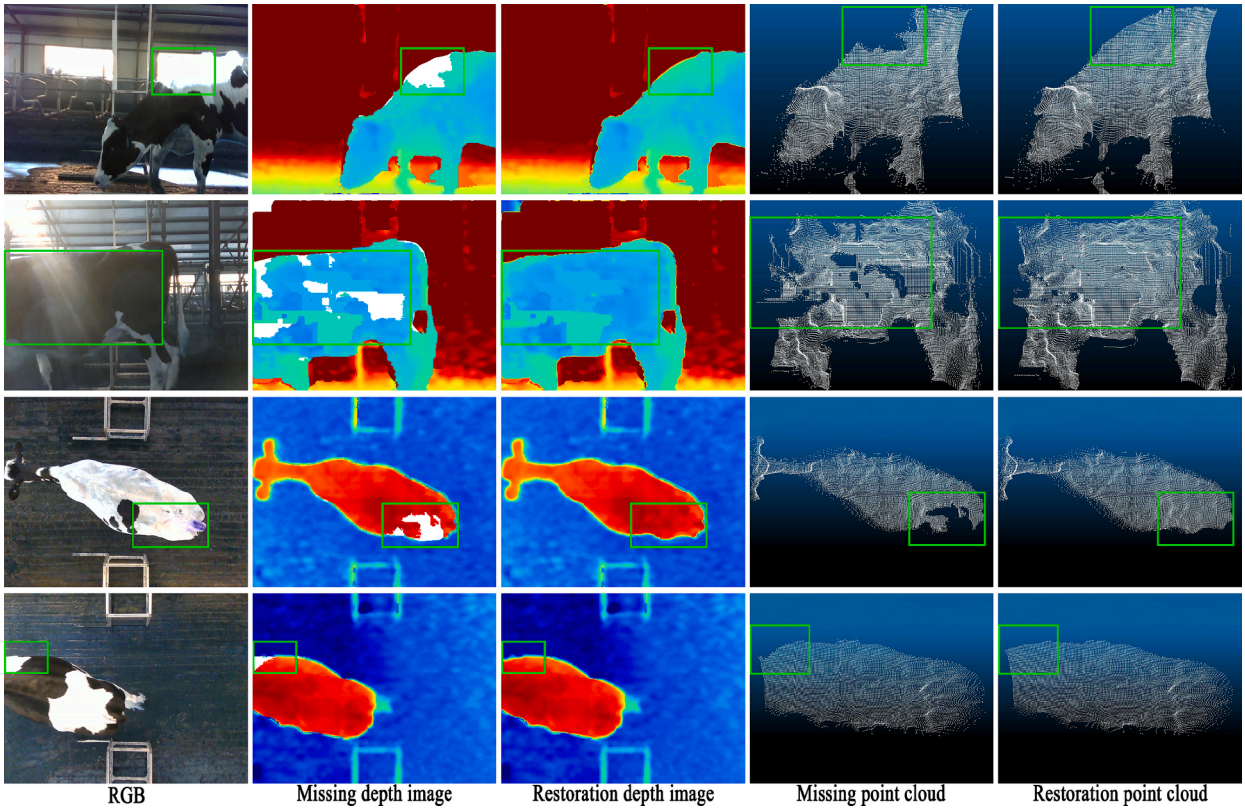



**Fig. 16.** The restoration results in Real Cow Depth Missing Dataset.

undulations and changes to improve the quality of depth image restoration.

## CRediT authorship contribution statement

**Yanxing Li:** Writing – review & editing, Writing – original draft, Methodology, Data curation, Conceptualization. **Xin Dai:** Writing – review & editing. **Baisheng Dai:** Writing – review & editing, Methodology, Funding acquisition, Formal analysis, Data curation. **Peng Song:** Validation, Resources. **Xinjie Wang:** Validation, Resources. **Xinchao Chen:** Validation, Software, Resources. **Yang Li:** Data curation. **Weizheng Shen:** Writing – review & editing, Supervision.

**Table 4**
Difference in evaluation indicators for cow body condition scoring at different depth images under zero error range.

| BCS | 0 error range | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Precision (%) | | | Recall rate (%) | | | F1-score (%) | | |
| | raw | restoration | diff | raw | restoration | diff | raw | restoration | diff |
| 2.5 | 0 | 10 | ↑10 | 0 | 14 | ↑14 | 0 | 12 | ↑12 |
| 2.75 | 5 | 13 | ↑8 | 5 | 5 | – | 5 | 7 | ↑2 |
| 3.0 | 26 | 29 | ↑3 | 17 | 52 | ↑35 | 21 | 37 | ↑16 |
| 3.25 | 21 | 36 | ↑15 | 18 | 19 | ↑1 | 19 | 25 | ↑6 |
| 3.5 | 16 | 29 | ↑13 | 19 | 29 | ↑10 | 18 | 29 | ↑11 |
| 3.75 | 8 | 11 | ↑3 | 14 | 10 | ↓4 | 10 | 11 | ↑1 |
| 4 | 0 | 11 | ↑11 | 0 | 5 | ↑5 | 0 | 7 | ↑7 |
| 4.25 | 15 | 19 | ↑4 | 19 | 33 | ↑14 | 17 | 24 | ↑7 |
| Average | 11 | 20 | ↑9 | 12 | 21 | ↑9 | 11 | 19 | ↑8 |

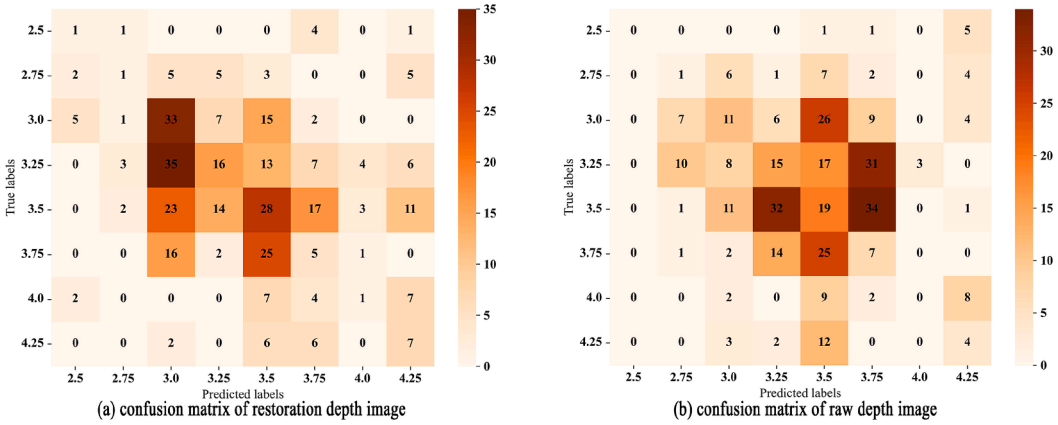



**Fig. 17.** Confusion matrix of scoring results from raw and restoration depth images.

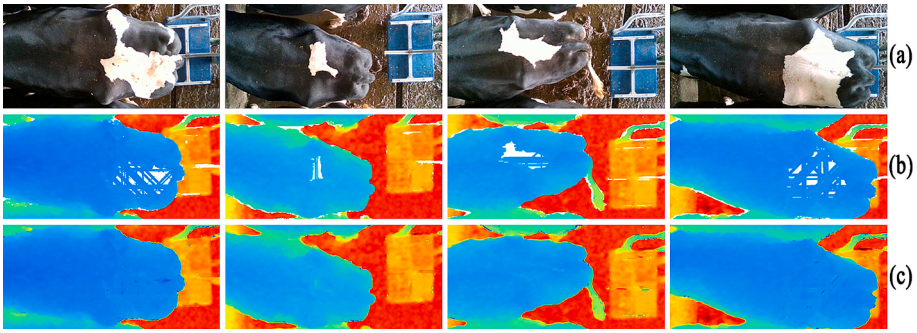



**Fig. 18.** The raw and restoration depth images for cow body condition scoring. (a) was the RGB image, (b) was the raw depth image, (c) was the restoration depth image.

## 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.

## Acknowledgments

This work was supported in part by the National Key Research and Development Program of China (Grant No. 2023YFD2000700), in part by the Key Research and Development Program of Heilongjiang Province (Grant No. 2022ZX01A24), in part by the National Natural Science Foundation of China (Grant No. 32072788), in part by the Modern Agricultural Industry Technology Collaborative Innovation Promotion System Construction Project of Heilongjiang Province (the Letter of Department of Agriculture and Rural Affairs of Heilongjiang Province (2021) No. 1492), and in part by the earmarked fund for CARS (Grant No. CARS-36).

## Data availability

Data will be made available on request.

## References

Atapour-Abarghouei, A., Breckon, T.P., 2017. Depthcomp: Real-time depth image completion based on prior semantic scene segmentation. In: 28th British Machine Vision Conference (BMVC). https://doi.org/10.5244/C.31.58.

Chodosh, N., Wang, C., Lucey, S., 2019. Deep convolutional compressed sensing for lidar depth completion. In: Computer Vision–ACCV 2018: 14th Asian Conference on Computer Vision, pp. 499–513. https://doi.org/10.1007/978-3-030-20887-5_31.

Dimitrievski, M., Veelaert, P., Philips, W., 2018. Learning morphological operators for depth completion. In: Advanced Concepts for Intelligent Vision Systems: 19th

International Conference, pp. 450–461. https://doi.org/10.1007/978-3-030-01449-0_38.
Eldesokey, A., Felsberg, M., Khan, F.S., 2018. Propagating confidences through cnns for sparse data regression. In: British Machine Vision Conference (BMVC), p. 14. https://doi.org/10.48550/arXiv.1805.11913.
Eldesokey, A., Felsberg, M., Holmquist, K., Persson, M.J.I., 2020. Uncertainty-aware cnns for depth completion: Uncertainty from beginning to end. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 12014–12023.
Gansbeke, W.V., Neven, D., Brabandere, B.D., Gool, L.V., 2019. Sparse and noisy lidar completion with rgb guidance and uncertainty. In: 2019 16th International Conference on Machine Vision Applications (MVA), pp. 1–6.
Garro, V., Mutto, C.D., Zanuttigh, P., Cortelazzo, G.M., 2009. A novel interpolation scheme for range data with side information. In: 2009 Conference for Visual Media Production, pp. 52–60. https://doi.org/10.1109/CVMP.2009.26.
He, C., Qiao, Y., Mao, R., Li, M., Wang, M., 2023. Enhanced litehrnet based sheep weight estimation using rgb-d images. Comput. Electron. Agric. 206, 107667. https://doi.org/10.1016/j.compag.2023.107667.
Hu, J.J., Bao, C.Y., Ozay, M., Fan, C.Y., Gao, Q., Liu, H.H., Lam, T.L., 2023. Deep depth completion from extremely sparse data: a survey. IEEE Trans. Pattern Anal. Mach. Intell. 45 (7), 8244–8264. https://doi.org/10.1109/tpami.2022.3229090.
Huang, Y.K., Wu, T.H., Liu, Y.C., Hsu, W.H., 2019. Indoor depth completion with boundary consistency and self-attention. In: Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops.
Islam, A.B.M.T., Scheel, C., Pajarola, R., Staadt, O., 2017. Robust enhancement of depth images from depth sensors. Comput. Graphics 68, 53–65. https://doi.org/10.1016/j.cag.2017.08.003.
Jaritz, M., Charette, R.D., Wirbel, E., Perrotton, X., Nashashibi, F., 2018. Sparse and dense data with cnns: Depth completion and semantic segmentation. In: 2018 International Conference on 3D Vision (3DV), pp. 52–60. https://doi.org/10.1109/3DV.2018.00017.
Jia, N., Kootstra, G., Koerkamp, P.G., Shi, Z., Du, S., 2021. Segmentation of body parts of cows in rgb-depth images based on template matching. Comput. Electron. Agric. 180, 105897. https://doi.org/10.1016/j.compag.2020.105897.
Li, J., Ma, W., Li, Q., Zhao, C., Tulpan, D., Yang, S., Ding, L., Gao, R., Yu, L., Wang, Z., 2022. Multi-view real-time acquisition and 3d reconstruction of point clouds for beef cattle. Comput. Electron. Agric. 197, 106987. https://doi.org/10.1016/j.compag.2022.106987.
Lu, K., Barnes, N., Anwar, S., Zheng, L., 2020. From depth what can you see? Depth completion via auxiliary image reconstruction. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 11306–11315.
Lu, S., Ren, X., Liu, F., 2014. Depth enhancement via low-rank matrix completion. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3390–3397.
Ma, F.C., Karaman, S., 2018. Sparse-to-dense: Depth prediction from sparse depth samples and a single image. In: 2018 IEEE International Conference on Robotics and Automation (ICRA), pp. 4796–4803. https://doi.org/10.1109/ICRA.2018.8460184.
Matsuo, T., Fukushima, N., Ishibashi, Y., 2013. Weighted joint bilateral filter with slope depth compensation filter for depth map refinement. In: International Conference on Computer Vision Theory and Applications, pp. 300–309. https://doi.org/10.5220/0004292203000309.
Park, J., Kim, H., Tai, Y.W., Brown, M.S., Kweon, I.S., 2014. High-quality depth map upsampling and completion for rgb-d cameras. IEEE Trans. Image Process. 23 (12), 5559–5572. https://doi.org/10.1109/TIP.2014.2361034.
Park, J., Joo, K., Hu, Z., Liu, C.K., So Kweon, I., 2020. Non-local spatial propagation network for depth completion. In: Computer Vision–ECCV 2020: 16th European Conference, pp. 120–136. https://doi.org/10.1007/978-3-030-58601-0_8.
Petschnigg, G., Szeliski, R., Agrawala, M., Cohen, M., Hoppe, H., Toyama, K., 2004. Digital photography with flash and no-flash image pairs. ACM Trans. Graphics 23, 664–672. https://doi.org/10.1145/1015706.1015777.
Ryu, K., Lee, K.I., Cho, J., Yoon, K.J., 2021. Scanline resolution-invariant depth completion using a single image and sparse lidar point cloud. IEEE Rob. Autom. Lett. 6 (4), 6961–6968. https://doi.org/10.1109/LRA.2021.3096499.
Senushkin, D., Romanov, M., Belikov, I., Patakin, N., Konushin, A., 2021. Decoder modulation for indoor depth completion. In: 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 2181–2188. https://doi.org/10.1109/IROS51168.2021.9636870.
Shen, W.Z., Zhang, Z., Dai, B.S., Wang, X.J., Zhao, K.X., Li, Y., 2023. Non-contact predicting method of dairy cow weight based on cow-detr and deep image. Trans. Chinese Society for Agric. Machinery 54 (08), 277–285. https://doi.org/10.6041/j.issn.1000-1298.2023.08.027.
Shi, W., Dai, B.S., Shen, W.Z., Sun, Y.K., Zhao, K.X., Zhang, Y.G., 2023. Automatic estimation of dairy cow body condition score based on attention-guided 3d point cloud feature extraction. Comput. Electron. Agric. 206, 107666. https://doi.org/10.1016/j.compag.2023.107666.
Sun, Y., Huo, P., Wang, Y., Cui, Z., Li, Y., Dai, B., Li, R., Zhang, Y., 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), 10140–10151. https://doi.org/10.3168/jds.2018-16164.
Tan, Z., Liu, J., Xiao, D., Liu, Y., Huang, Y., 2023. Dual-stream fusion network with convnextv2 for pig weight estimation using rgb-d data in aisles. Animals 13 (24), 3755. https://doi.org/10.3390/ani13243755.
Uhrig, J., Schneider, N., Schneider, L., Franke, U., Brox, T., Geiger, A., 2017. Sparsity invariant cnns. In: 2017 International Conference on 3D Vision (3DV), pp. 11–20. https://doi.org/10.1109/3DV.2017.00012.
Wang, H., Wang, M., Che, Z., Xu, Z., Qiao, X., Qi, M., Feng, F., Tang, J., 2022. Rgb-depth fusion gan for indoor depth completion. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 6209–6218.
Winkler, Z., Boucheron, L.E., Utsumi, S., Nyamuryekung'e, S., Mcintosh, M., Estell, R.E., 2024. Effects of dataset curation on body condition score (bcs) determination with a vision transformer (vit) applied to rgb+depth images. Smart Agric. Technol. 8, 100482. https://doi.org/10.1016/j.atech.2024.100482.
Xie, Z., Yu, X., Gao, X., Li, K., Shen, S., 2024. Recent advances in conventional and deep learning-based depth completion: a survey. IEEE Trans. Neural Networks Learn. Syst. 35 (3), 3395–3415. https://doi.org/10.1109/TNNLS.2022.3201534.
Yang, G., Xu, X., Song, L., Zhang, Q., Duan, Y., Song, H., 2022. Automated measurement of dairy cows body size via 3d point cloud data analysis. Comput. Electron. Agric. 200, 107218. https://doi.org/10.1016/j.compag.2022.107218.
Yu, Q., Chu, L., Wu, Q., Pei, L., 2021. Grayscale and normal guided depth completion with a low-cost lidar. In: 2021 IEEE International Conference on Image Processing (ICIP), pp. 979–983. https://doi.org/10.1109/ICIP42928.2021.9506577.
Yu, J., Lin, Z., Yang, J., Shen, X., Lu, X., Huang, T.S., 2019. Free-form image inpainting with gated convolution. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp. 4471–4480.
Zhang, Y., Guo, X., Poggi, M., Zhu, Z., Huang, G., Mattoccia, S., 2023. Completionformer: depth completion with convolutions and vision transformers. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 18527–18536.