Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

GoatPoinTr: Single-View Point Cloud Completion and Keypoint Detection Network for Goat Body Measurement

Bo Jin
NORTHWEST A&F UNIVERSITY
Guorui Wang
NORTHWEST A&F UNIVERSITY
Chong He
NORTHWEST A&F UNIVERSITY
Jingze Feng
NORTHWEST A&F UNIVERSITY
Yongliang Qiao
University of Adelaide
Fengtao Nan
NORTHWEST A&F UNIVERSITY
Meili Wang
wml@nwsuaf.edu.cn

NORTHWEST A&F UNIVERSITY

Research Article

Keywords: Point cloud completion, Livestock body measurement, Keypoint detection, Precision farming

Posted Date: May 18th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9606991/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

CGI2026 manuscript No.
(will be inserted by the editor)

GoatPoinTr: Single-View Point Cloud Completion and Keypoint Detection Network for Goat Body Measurement

Bo Jin[1] · **Guorui Wang**[1] · **Chong He**[1] · **Jingze Feng**[1] · **Yongliang Qiao**[5] · **Fengtao Nan**[1] · **Meili Wang**[1,2,3,4*]

Abstract Automatic non-contact body measurement of livestock is crucial for precision farming, yet existing methods rely on complex multi-view camera systems or suffer from incomplete point cloud representation and inaccurate keypoint detection from single-view acquisitions. This paper presents GoatPoinTr, a joint point cloud completion and keypoint detection network built upon AdaPoinTr for single-view goat body measurement. An online back-projecting masking strategy with viewpoint jittering is proposed to realistically simulate single-view acquisition during training by projecting complete point clouds onto virtual camera planes and retaining only visible surface points. Outlook Attention is introduced to improve fine-grained geometric detail recovery, particularly for thin structures such as limbs. A keypoint detection head is further integrated into the network through a cross-optimization mechanism that enables bidirectional refinement between point cloud completion and body measurement keypoint prediction within each forward pass. A goat point cloud benchmark comprising 5,712 samples from 82 goats with 10 manually annotated body measurement keypoints is established to support point cloud completion and body size estimation research. Extensive experiments demonstrate that GoatPoinTr achieves CD-l_1 of 9.326 and CD-l_2 of 0.264, reducing CD-l_1 by 13.4% and CD-l_2 by 33.3% compared to AdaPoinTr. The mean absolute percentage errors across seven body dimensions range from 2.66% to 5.55%, comparable to multi-view methods and meeting livestock measurement accuracy standards.

Keywords Point cloud completion · Livestock body measurement · Keypoint detection · Precision farming

✉ Meili Wang
wml@nwsuaf.edu.cn
[1] College of Information Engineering, Northwest A&F University, Yangling 712100, China
[2] Key Laboratory of Agricultural Internet of Things, Ministry of Agriculture, Yangling 712100, China
[3] Shaanxi Key Laboratory of Agricultural Information Perception and Intelligent Service, Yangling 712100, China
[4] Shaanxi Engineering Research Center of Agricultural Information Intelligent Perception and Analysis, Yangling 712100, China
[5] Australian Institute for Machine Learning (AIML), The University of Adelaide, 5005, Australia

1 INTRODUCTION

Accurate livestock body measurements are essential for precision livestock farming, as they provide critical data for breed selection, health monitoring, feed optimization, and economic valuation [1]. Traditional measurement methods, such as using measuring rods, tapes, and calipers, are labor-intensive, time-consuming, and can cause stress or injury to animals. With advancements in computer vision and 3D sensing technologies, non-contact measurement methods based on 2D images [2, 3, 4] and 3D point clouds [5, 6, 7, 8] have gained increasing popularity in livestock farming, enabling automated body size estimation while ensuring animal welfare.

Among these, 3D point cloud-based methods offer distinct advantages in directly capturing spatial geometry, enabling precise measurement of three-dimensional body traits. Traditional 3D approaches rely on multi-view depth camera systems, which capture the animal from multiple perspectives and register the resulting point clouds into a unified representation [5, 9, 7, 10]. While these systems can produce complete point clouds, they require complex hardware setup, careful calibration, and temporal synchronization across cameras. These demanding requirements make multi-view systems impractical for routine use in real-world farming environments, particularly in open pastures or small-scale operations.

To address these practical challenges, researchers have explored single-view alternatives that capture data from a single depth camera, offering a simpler and more cost-effective solution [11,12,8]. However, single-view acquisition inherently produces incomplete point clouds due to self-occlusion, which can lead to substantial measurement errors, especially for body dimensions that require information from occluded regions. Some studies have attempted to address this incompleteness through symmetry-based completion [13,14], but these methods are sensitive to posture variations and require precise body alignment, limiting their applicability to freely moving animals.

Deep learning-based point cloud completion offers a more robust alternative, capable of recovering complete 3D shapes from partial observations without relying on geometric assumptions. Methods such as PCN [2], SnowflakeNet [15], SeedFormer [16], PoinTr [17], and AdaPoinTr [18] have demonstrated remarkable capability in reconstructing complete shapes on synthetic benchmarks. However, these methods are primarily designed for rigid objects and have not been explored for livestock body reconstruction, where the target shapes are deformable, articulated, and captured under varying postures. Furthermore, existing completion methods focus solely on shape recovery and lack the ability to simultaneously detect anatomically meaningful keypoints required for body measurement. When completion and keypoint detection are performed as separate sequential stages, the two tasks cannot benefit from each other, resulting in suboptimal performance for both.

A unified framework that jointly performs point cloud completion and keypoint detection from single-view inputs would not only improve measurement accuracy through mutual task reinforcement but also simplify the overall processing pipeline by eliminating the need for separate sequential stages. Such a framework, however, must address the realistic simulation of single-view occlusion patterns during training, the recovery of fine-grained geometric details in structurally thin regions such as limbs and joints, and the effective interaction between completion and keypoint detection to ensure that both tasks benefit from shared feature representations.

Based on the above analysis, this paper proposes GoatPoinTr, a joint point cloud completion and keypoint detection network built upon AdaPoinTr for single-view goat body measurement. It consists of three collaboratively working components: an online back-projecting masking strategy with viewpoint jittering, Outlook Attention, and a keypoint detection head with a cross-optimization mechanism. The back-projecting masking strategy projects complete point clouds onto virtual camera planes and retains only visible surface points, realistically simulating single-view acquisition during training. Unlike the crop masking in AdaPoinTr, which randomly removes the farthest points and may not reflect actual occlusion patterns, our method faithfully reproduces the visibility characteristics of a single depth camera. Additionally, viewpoint jittering is applied during training to prevent overfitting to specific viewing angles. Outlook Attention [19] is introduced to replace standard self-attention in the Transformer encoder. By generating attention weights from local feature representations rather than global pairwise comparisons, Outlook Attention captures fine-grained geometric patterns in structurally complex regions such as limbs and joints, which are critical for accurate body measurement. The keypoint detection head is integrated into the network through a cross-optimization mechanism that splits global features into completion and keypoint branches, enabling bidirectional refinement within each forward pass. The coarse point features inform keypoint prediction, while the detected keypoints feed back to refine the completion process, creating a mutually beneficial loop that improves both tasks simultaneously.Using the detected keypoints and completed point cloud, seven body dimensions are computed, including body length (BL), body height (BH), chest width (CW), chest girth (CG), chest depth (CD), hip height (HH), and hip width (HW).

The principal innovations and contributions of this work are summarized as follows:
(1) We proposed GoatPoinTr, a joint point cloud completion and keypoint detection network for single-view goat body measurement, capable of simultaneously recovering complete point clouds and detecting 10 body measurement keypoints.
(2) We designed an online back-projecting masking strategy with viewpoint jittering that realistically simulates single-view acquisition during training, eliminating the need for pre-rendered partial point clouds.
(3) We introduced Outlook Attention into the 3D Transformer encoder for capturing fine-grained geometric details in complex regions, while developing a keypoint detection head with a cross-optimization mechanism for bidirectional refinement between point cloud completion and keypoint detection.
(4) We established a goat point cloud benchmark comprising 5,712 samples from 82 goats with 10 manually annotated body measurement keypoints.

2 RELATED WORK

Livestock body measurement based on 3D point clouds can be broadly categorized into two paradigms: (1) multi-

view acquisition systems that construct complete point clouds from multiple depth cameras, and (2) single-view approaches that work directly with partial point clouds or employ completion techniques to recover the full body shape. In addition, as our method builds upon deep learning-based point cloud completion, we review the relevant progress in this area.

2.1 Multi-view Livestock Body Measurement

Several studies have demonstrated the feasibility of using multi-view depth cameras for livestock body size measurement. [5] used six Microsoft Kinect depth cameras to measure phenotypic traits of dairy cattle, constructing complete body surfaces from multiple perspectives. Similarly, [9] developed a portable multi-camera system and detected keypoints based on the shape constraints of pig legs. [7] introduced the PointNet++ model for segmenting pig point clouds, followed by keypoint detection and body measurements using point cloud slicing techniques. [20] used the orb calibration method [21] to acquire external references for multi-view camera registration prior to data acquisition. [10] proposed a PointStack-based framework for automatic 3D goat body measurement using multi-view point clouds with keypoint detection.[6] employed three depth cameras combined with non-rigid 3D shape recovery to achieve accurate body measurements of live cattle.Despite their effectiveness, these multi-view systems require complex hardware setup, precise calibration, and temporal synchronization across cameras. Moreover, the fixed camera configurations restrict measurements to dedicated channels or enclosures, limiting routine deployment in real-world farming environments.

2.2 Single-view Livestock Body Measurement

Given the practical challenges of multi-view systems, researchers have explored single-view alternatives that capture data from a single depth camera, offering a simpler and more cost-effective solution. [11] captured single-view point clouds of cattle using a LiDAR sensor and applied Iterative Closest Point (ICP) and Greedy Projection Triangulation methods for body parameter estimation. [12] collected single-view point clouds of dairy cow backs and applied PCA for automatic body measurement extraction. [8] estimated cattle body weight using an improved PointNet++ segmentation model based on lateral single-view point clouds. [14] proposed a point cloud reparation method using moving least squares sampling for extracting body sizes of beef cattle from unilateral point clouds. [13] used a symmetrical plane method to mirror and complete the single-view body representation for pig body size measurement. While these single-view methods significantly reduce hardware complexity, they face a fundamental limitation: the captured point clouds are inherently incomplete due to self-occlusion, which restricts the measurable body dimensions. To address this incompleteness, symmetry-based methods [13, 14] reconstruct occluded regions by mirroring visible points along the body's symmetrical plane. However, these approaches require precise posture alignment and are sensitive to body orientation variations. Deep learning-based point cloud completion presents a more robust alternative without relying on geometric assumptions.

2.3 Point Cloud Completion

Point cloud completion aims to recover a complete 3D shape from partial observations, which is a fundamental problem in 3D computer vision [22]. Early learning-based approaches such as PCN [2] employ an encoder-decoder framework with a coarse-to-fine generation strategy to produce complete point clouds. TopNet [23] introduces a hierarchical rooted tree structure for structured point cloud decoding. GRNet [24] leverages 3D grids as an intermediate representation to regularize the output and improve completion quality.

More recently, Transformer-based methods have achieved substantial improvements. PoinTr [17] reformulates point cloud completion as a set-to-set translation problem using geometry-aware Transformers, enabling better structural knowledge learning and detail preservation. AdaPoinTr [18] further introduces an adaptive query generation mechanism and a novel denoising task, achieving state-of-the-art performance on standard benchmarks including PCN, ShapeNet-55, and ShapeNet-34. SnowflakeNet [15] proposes a snowflake point deconvolution module with skip-transformer layers for progressive point generation, effectively capturing both global structure and local details. SeedFormer [16] introduces a novel Patch Seeds representation that preserves regional information of local patterns, combined with an Upsample Transformer for high-fidelity coarse-to-fine completion. ProxyFormer [25] separates existing and missing parts of point clouds and designs a missing part sensitive transformer with proxy alignment for efficient and accurate completion. PointAttN [26] presents a fully attention-based architecture that enhances point cloud completion through pure attention mechanisms without convolution.

Despite these significant advances, existing point cloud completion methods are primarily developed and evaluated on synthetic object datasets (e.g., ShapeNet,

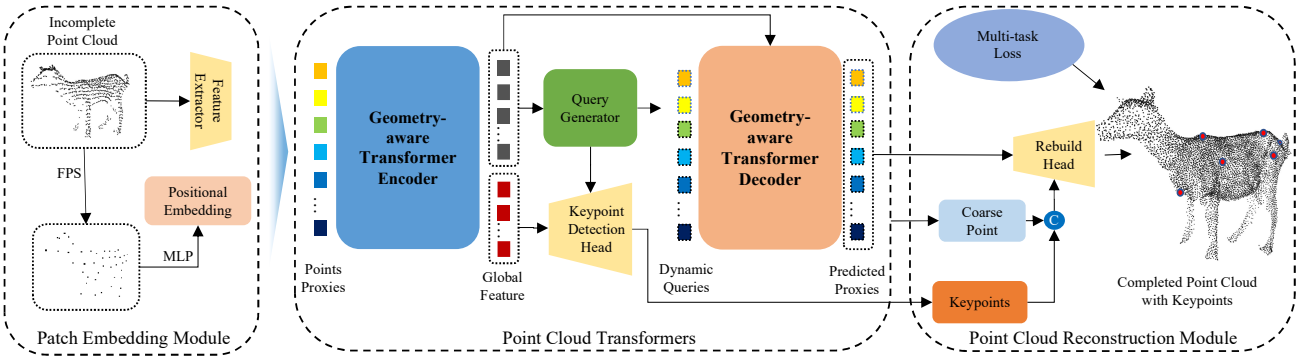


Fig. 1 GoatPoinTr point cloud completion-based model architecture

Completion3D) and have rarely been applied to livestock body reconstruction, where the target shapes are deformable, articulated, and captured under varying postures. In this work, we propose GoatPoinTr, which builds upon AdaPoinTr and extends it with joint keypoint detection to address the specific challenges of single-view goat point cloud completion and body measurement.

3 METHOD

3.1 GoatPoinTr Model

We propose GoatPoinTr, a joint point cloud completion and keypoint detection network built upon AdaPoinTr. GoatPoinTr introduces three key modifications: (1) an online back-projecting masking strategy with viewpoint jittering that realistically simulates single-view point cloud acquisition; (2) Outlook Attention that replaces self-attention in the Transformer encoder for enhanced fine-grained geometric detail recovery; and (3) a keypoint detection head with a cross-optimization mechanism that enables bidirectional refinement between point cloud completion and keypoint detection. The overall network architecture is shown in Figure 1.

3.1.1 Back-projecting Masking Strategy

Point cloud completion aims to reconstruct a full point cloud from a partial representation. Existing methods like PCN [2] and MAE3D [27] rely on effectively utilizing the available partial data. A critical component of this is the masking strategy, which should retain enough information for training.

The crop masking strategy in AdaPoinTr randomly removes the farthest n points from a selected viewpoint, which may not reflect real-world single-view scenarios. For instance, when observing a goat from the side, if the legs are separated, the method may crop out the points of the far-side legs, missing essential information.

To address this, we propose a novel online back-projecting masking strategy with viewpoint jittering to avoid overfitting. This method is more flexible and efficient compared to offline back-projecting approaches like PCN [2], Completion3D [23], and MVP [28].

We first project the point cloud $\mathcal{P}$ on to a plane with a normal vector $\mathbf{n} = v - o$ at viewpoint v, obtaining $\mathcal{P}'$, where o is the origin. This can be expressed as:

$$\mathcal{P}' = \mathcal{P} - \left(\frac{(\mathcal{P} - v) \cdot \mathbf{n}}{\|\mathbf{n}\|^2} \right) \mathbf{n}. \tag{1}$$

We then establish a bijection between $\mathcal{P}$ and $\mathcal{P}'$, denotes as $\mathcal{P} \leftrightarrow \mathcal{P}'$. Subsequently, we divide $\mathcal{P}'$ into grids. Within each grid cell on the plane of $\mathcal{P}'$, denoted as p'_i, are sorted according to their distance from the corresponding original points in $\mathcal{P}$. From this sorted list, we select the top k points. This procedure can be mathematically formulated as:

$$\mathcal{P}'_{\text{plane}} = \bigcup_j \left\{ p'_i \in G_j \mid i \in \text{Top}_K \left(\text{d}_{G_j} \right) \right\} \tag{2}$$

where $\mathcal{P}'_{\text{plane}}$ represents the collection of top k points across all grid cells, G_j denotes the j-th grid cell, and d_{G_j} is the distance of points in grid cell G_j to the original points in $\mathcal{P}$. And then we can index to the original point $\mathcal{P}$ with the mapping function $\mathcal{P}' \to \mathcal{P}$.

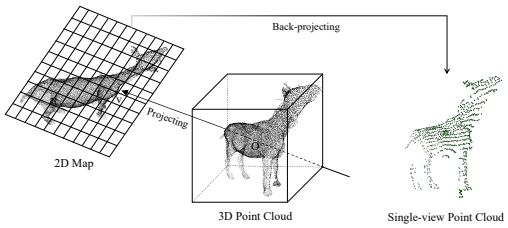


Fig. 2 Back projecting masking strategy scheme

As shown in Figure 2, this process projects and discretizes the input point cloud into a 2D map. It divides the point set into grids and sorts them based on the distance from the original points to the projection plane. It then maps them back into 3D space, retaining the "front view" point cloud. This effectively simulates a scenario where the scene is captured from a single-viewpoint. In addition, the viewpoint is jittered online during the training process to prevent overfitting.

3.1.2 Outlook Attention for Fine-Grained Detail Recovery

The original AdaPoinTr encoder employs standard self-attention in its geometry-aware Transformer blocks, which computes global pairwise attention weights across all point proxy tokens. While effective for capturing long-range dependencies, self-attention operates at a coarse level and may dilute focus on local geometric details critical for livestock body reconstruction, particularly in thin, articulated structures such as legs and ears.

Under the proposed back-projecting masking strategy, the model is required to complete the occluded side of the goat body from visible surface points. This task demands strong local feature extraction to fully exploit the geometric information present in the visible point cloud. Inspired by the VOLO framework [19], we replace the self-attention layers in the geometry-aware Transformer encoder with Outlook Attention to enhance the model's capacity for fine-grained detail recovery.

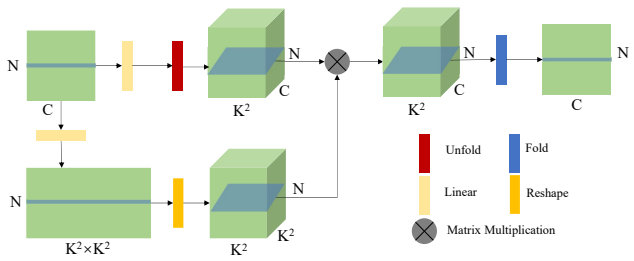


Fig. 3 Illustration of outlook attention

As shown in Figure 3, for each input point proxy token of feature dimension C, the local neighborhood is defined by selecting its K^2 nearest neighbors in 3D Euclidean space via KNN search, where K^2 denotes the neighborhood size analogous to the $K \times K$ local window in the original image-based VOLO formulation. Outlook Attention then operates through two parallel branches to aggregate fine-grained local context.

In the attention branch, the feature of each anchor token (size $N \times C$) is projected through a linear layer and reshaped to produce a local attention weight matrix of size $N \times K^2 \times K^2$, which encodes pairwise attention weights among the K^2 neighboring tokens within the local neighborhood, without requiring the expensive dot-product computation between all token pairs used in standard self-attention. In the value branch, the input features (size $N \times C$) are first projected via a linear layer to obtain a dense value representation. The value features of the K^2 nearest neighboring tokens for each anchor are then gathered via KNN indexing, analogous to the unfold operation in the original image-based formulation, yielding a local value tensor of size $N \times K^2 \times C$.

The attention weights from the attention branch are passed through a softmax normalization and applied to the corresponding local value tensor via matrix multiplication, producing locally-attended output features of size $N \times K^2 \times C$. These are subsequently aggregated back to the original token resolution through a fold operation, yielding the final output of size $N \times C$ that encodes fine-grained spatial context from neighboring proxy tokens.

Compared to standard self-attention, this locally-informed mechanism focuses computational resources on neighboring geometric structures, improving the recovery of detailed body contours such as limbs and joints, as validated in our ablation study (Section 4.5, Model C).

3.1.3 Keypoints detection head

Goat body measurement points are distributed across the body. Complete point cloud data allows easy identification of these points, while the measurement points themselves provide positional information on the goat's surface. Leveraging GoatPoinTr's capability in point cloud completion, we developed a keypoint detection head for accurate keypoint identification prior to body size measurement. Additionally, we introduced a cross-mechanism that iteratively optimizes both point cloud completion and keypoint detection within the feed-forward network. The design philosophy and implementation details of the proposed keypoint detection are as follows.

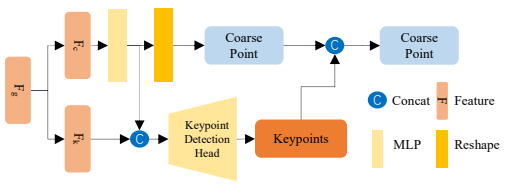


Fig. 4 Illustration of keypoints detection

To leverage the global features from the Transformer in AdaPoinTr for both point cloud completion and keypoint detection, we split the global features $\mathbf{F}_g$ into two parts: coarse feature $\mathbf{F}_c$ for point completion and keypoint feature $\mathbf{F}_k$ for detection:

$$\mathbf{F}_c, \mathbf{F}_k = split(\mathbf{F}_g, (d_c, d_k)), \tag{3}$$

where d_c and d_k denote the dim of $\mathbf{F}_c$ and $\mathbf{F}_k$.

The final keypoint features $\mathbf{F}_{fk}$ are updated by combining two sets of features: one set from the coarse points processed by a learnable MLP and the other set of keypoint features derived from the global features. The formula for this process can be expressed as follows:

$$\mathbf{F}_{fk} = Concat\left(\mathrm{MLP}\left(\mathbf{F}_k\right), \mathrm{MLP}\left(\mathcal{P}_c\right)\right). \tag{4}$$

This step merges features from both sources to refine keypoint detection accuracy. After normalization and activation, the final 10 keypoints are obtained and used for precise goat body measurement mapping.

Additionally, the detected keypoints are fed back to refine the coarse points (Figure 4). This bidirectional process improves point cloud completion by leveraging keypoint detection accuracy, while keypoint detection benefits from the completion results. Through this cross-optimization mechanism, the model evolves iteratively, achieving progressively more accurate point cloud completion and keypoint detection.

3.1.4 Multi-task Loss Function

In this paper, a multi-task joint learning strategy is used to simultaneously guiding the prediction of the coarse point, the complementary point cloud, and the keypoint of body measurement. The reconstructed point cloud effectiveness is measured by Chamfer Distance (CD), defined as:

$$\mathrm{CD}(\mathcal{P}, \mathcal{G}) = \frac{1}{|\mathcal{P}|}\sum_{p\in\mathcal{P}}\min_{g\in\mathcal{G}}\|p-g\| + \frac{1}{|\mathcal{G}|}\sum_{g\in\mathcal{G}}\min_{p\in\mathcal{P}}\|g-p\|. \tag{5}$$

For keypoint detection, we use Mean Square Error (MSE), defined as:

$$\mathrm{MSE}(P_k, P_q) = \frac{1}{n}\sum_{i=1}^{n}(P_{k,i} - P_{q,i})^2, \tag{6}$$

where P_k, P_q are the predicted and true keypoint locations, respectively, and n is the number of keypoints.

To optimize both point cloud completion and keypoint detection, we define the multi-task joint loss as:

$$J = \alpha J_{PC} + \beta J_{MSE}, \tag{7}$$

where J_{PC} follows AdaPoinTr [18], and J_{MSE} is from Equation 6. The hyperparameters α and β control the weight of each term.

3.2 Body Measurement

In previous sections, we completed the goat point cloud completion and keypoint detection tasks, from which we now calculate the goat body size, as shown in Figure 5. We defined a set of keypoints p1, p2, ..., p10, with their specific definitions provided in 4.1.2.

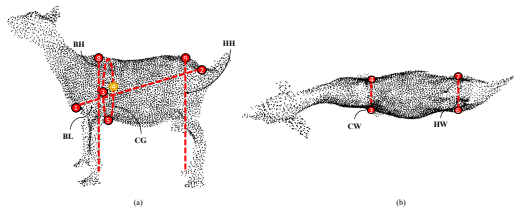


Fig. 5 Schematic of point cloud based goat body size measurement (a. side-view, b. top-view)

As described in 4.1.1, we employ a unified coordinate system. Using the z-coordinates of p_8 (x_8, y_8, z_8) and p_9 (x_9, y_9, z_9), we calculate the body height (BH) and hip height (HH):

$$\mathrm{BH} = z_8, \quad \mathrm{HH} = z_9. \tag{8}$$

The Euclidean distance is used to measure length indices such as body length (BL) and chest depth (CD), with body length calculated as the distance between the outer point of the iliac bone (p_1) and the posterior internal extremity of the ischium (p_2), and chest depth as the distance from the highest point of the chest (p_{10}) to the lowest point (p_5):

$$\mathrm{BL} = \|p_1 - p_2\|, \mathrm{CD} = \|p_{10} - p_5\|. \tag{9}$$

Distance between the posterior corner points of the shoulder (p_3 and p_4) is measured to calculate the chest width (CW), and distance between the hip joint points perpendicular to the back (p_6 and p_7) is measured to calculate the hip width (HW):

$$\mathrm{CW} = \|p_3 - p_4\|, \mathrm{HW} = \|p_6 - p_7\|. \tag{10}$$

For calculating the chest girth (CG), we can determine the plane in which the chest is located by p_3, p_4 and the lowest point of the chest (p_5). Then the normal vector of the plane is $\mathbf{n} = (p_3 - p_5) \times (p_4 - p_5)$. We collect points from the point cloud that are within a certain threshold distance δ from the slicing plane: $\mathcal{P}_{slice} = \{p_i \in \mathcal{P} \mid \left|(p_i - p_5) \cdot \frac{\mathbf{n}}{||\mathbf{n}||}\right| \leq \delta\}$. However, the

sliced point cloud will be affected by the point cloud of the goat legs as shown in Figure 5. Since we have a priori obtained the lowest point of the chest (p_5), we can remove the points in the slice with z-coordinates lower than $p_5(x_5, y_5, z_5)$:

$$\mathcal{P}_{\text{CG}} = \{p_i \in \mathcal{P}_{\text{slice}} \mid z_i \geq z_5\}. \tag{11}$$

Inspired by other previous work [10], the chest slice point cloud is further sorted using an optimized traveling salesman problem (TSP) algorithm. The sorted point cloud slice is defined as $\mathcal{P}_{\text{CG}} = \{c_1, c_2, \ldots, c_n\}$, and CG can be calculated through $\mathcal{P}_{\text{CG}}$:

$$\text{CG} = \|c_1 - c_n\|_2 + \sum_{i=1}^{n-1} \|c_i - c_{i+1}\|_2. \tag{12}$$

4 EXPERIMENT

4.1 Dataset

4.1.1 Data Acquisition and Preprocessing

The ground truth point cloud data were collected at a livestock teaching and experimental station of Northwest A&F University. As shown in Figure 6, a dedicated channel was set up within the goats' activity area. Three synchronized Azure Kinect DK depth cameras were positioned above and on both sides of the channel. To obtain the most complete ground truth point cloud, 2mm diameter wires were installed along the sides of the channel instead of railings, ensuring minimal impact on the point cloud after farthest point sampling. Intrinsic and extrinsic parameters of the cameras were calibrated using [29].

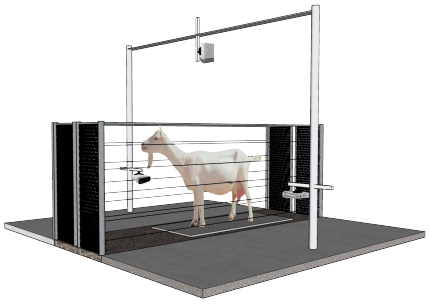


Fig. 6 Schematic diagram of the data acquisition equipment

The point clouds from each viewpoint were merged into a complete point cloud using extrinsic parameters. RANSAC [30] was applied for ground removal, and DBSCAN [31] was used to extract the goat point clouds from the complex background, followed by denoising to obtain smooth, high-quality data.

Next, the ground normal $\mathbf{z}$ from RANSAC segmentation and the body orientation $\mathbf{x}$ of the goat point cloud, calculated using PCA [32], are used as two axes. The cross product $\mathbf{y} = \mathbf{x} \times \mathbf{z}$ is computed to determine the y-axis. The coordinate origin is set at the centroid of the point cloud, establishing a unified coordinate system for further processing.

Since there is no dedicated public point cloud dataset for goats, all data were collected by us. In total, 82 goats were recorded, resulting in 357 point cloud datasets. The goats' ages ranged from 6 to 28 months, and various postures during movement were captured. For data augmentation, the point clouds were rotated at 8 angles (0°, 45°, 90°, 135°, 180°, 225°, 270°, 315°) and sampled twice using Farthest Point Sampling [33], generating a total of 5,712 goat point cloud datasets.

4.1.2 Keypoint Definition

In the GoatPoinTr keypoint detection task, the model relies on manually annotated keypoints. To minimize annotation variability caused by posture, a single trained annotator performed all annotations following expert-defined guidelines.

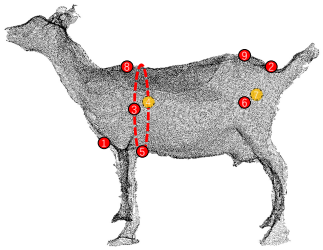


Fig. 7 Schematic representation of keypoints in goat body measurement (yellow indicates keypoints obscured from view)

As shown in Figure 7, we defined 10 keypoints for body size measurement: the outer point of the iliac bone (p_1), the posterior internal extremity point of the ischium (p_2), the posterior corner points of the shoulder (p_3 and p_4), the lowest point of the chest (p_5), the hip joint points perpendicular to the back (p_6 and p_7), the highest point of the withers (p_8), the highest point of the hip (p_9), and the highest point of the chest (p_{10}).

4.1.3 Data Splitting

To prevent data leakage, the dataset was split at the individual goat level: 57 goats (70%) for training, 8 goats (10%) for validation, and 17 goats (20%) for testing. No

goat appears in more than one split. Data augmentation was applied only to the training set. Hyperparameters including loss weights α and β, grid resolution, and learning rate schedule were tuned on the validation set. Reported results are evaluated on the held-out test set.

4.2 Experimental Setup

To ensure comprehensive body visibility for measurement, we positioned the evaluation viewpoint at a 45° angle on both sides of the goat. During training, the best-performing model parameters on the validation set were saved, and the model was subsequently applied to the test set for evaluation. Since there are no available supervised keypoint detection methods in point cloud learning, keypoint detection experiments are presented in the ablation study (Section 4.5). Finally, we performed automated body size measurements and compared them with manual measurements using measuring sticks and tape, as well as with the multi-view approach described in Section 4.6.

We used PyTorch [34] to train the model on a server equipped with an Intel Xeon Platinum 8160T CPU @ 2.10 GHz, 128 GB RAM, and a 24 GB NVIDIA RTX A5000 GPU, running in a CUDA 11.7.0 parallel computing environment. The model was trained for 300 epochs using the AdamW optimizer with an initial learning rate of 1×10^{-4} and a cosine annealing schedule.

4.3 Evaluation Metrics

In assessing the performance of GoatPoinTr for point cloud completion, we adopt the Chamfer Distance (CD, Equation 5) as our primary evaluation metric, in line with previous studies [2,24,27]. Consistent with the methodology of AdaPoinTr [18], we use two variants of CD: CD-l_1 (L1-norm) and CD-l_2 (L2-norm). Additionally, we use the F-Score with a threshold of 1% to evaluate the accuracy of point cloud completion, based on precision and recall.

For the keypoint detection task, following the practices in 3D human pose estimation [35], we use the Mean Per Joint Position Error (MPJPE) and Percentage of Correct Keypoints (PCK) as evaluation metrics.

Finally, to validate the accuracy of body measurements, we use Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) as metrics.

4.4 Goat Point Cloud Completion Results

We report the performance of our models in Table 1 and compare them with existing models, which we implemented on benchmarks based on their open-source code for FoldingNet [36], PCN [2], TopNet [23], GRNet [24], PoinTr [17], and AdaPoinTr [18]. We implemented these models using the best hyperparameters reported in their original papers. To simulate single-view scenarios, we applied the online cropping method from AdaPoinTr along with our online back-projecting strategy to mask the complete point clouds.

As shown in Figure 8, FoldingNet produces a manifold shape that does not recover the goat body, while PCN and TopNet struggle with vague head and leg shapes. PCN exhibits more patch-wise representation, and TopNet produces a more diffuse point cloud. GRNet fails to reconstruct the well-defined goat body, with significant noise points throughout. The PoinTr series offers a clearer body shape, but its head and legs remain sparse and coarse. AdaPoinTr and our method both recover a better overall goat body shape. However, AdaPoinTr remains inhomogeneous, lacking accuracy in representation of the legs. In contrast, our method achieves more accurate detail in these areas.

Our back-projecting masking strategy (Figure 9) reveals more leg details than the cropping strategy (Figure 8). Methods other than the PoinTr series exhibit limitations in Figure 9 similar to those observed in Figure 8. Due to the lack of an adaptive query generation mechanism, PoinTr does not produce uniformly distributed points across different body parts.

Under the back-projecting masking strategy, although AdaPoinTr's self-attention is limited in capturing fine details, its contours—particularly around the legs—are closer to the ground truth than those produced under random cropping. Our method, leveraging outlook attention, achieves finer limb details closer to GT, as shown in Table 1.

As shown in Table 1, our back-projecting strategy significantly reduces CD across all methods compared to AdaPoinTr's cropping. This improvement stems from its ability to better capture single-view goat acquisition information, as confirmed by the figure, demonstrating its superior simulation of real-world scenarios.

Meanwhile, our model achieves the best results in terms of F-Score and CD under the back-projecting strategy. As shown in the Figure 8 and Figure 9, we display some completion results of all methods and find that our method performs better. Other methods only coarsely complete shapes with less-than-ideal geometric details, while our method still manages to complete point clouds with higher fidelity.

Table 1 Completion performace of different backbone on the goat dataset (CD-l_1, CD-l_2 in units of 1e-3)

	Crop			**Back-projecting**		
	F-Score@1% ↑	CD-l_1 ↓	CD-l_2 ↓	F-Score@1% ↑	CD-l_1 ↓	CD-l_2 ↓
FoldingNet	0.188	25.548	2.268	0.191	23.905	1.876
PCN	0.227	21.805	1.652	0.246	19.318	1.144
TopNet	0.251	18.666	1.166	0.242	17.888	0.920
GRNet	0.302	20.229	1.686	0.281	17.945	0.981
PoinTr	0.369	15.423	0.963	0.337	15.565	0.832
AdaPoinTr	0.584	12.585	0.906	0.555	10.768	0.396
Ours	**0.596**	**11.374**	**0.763**	**0.659**	**9.326**	**0.264**

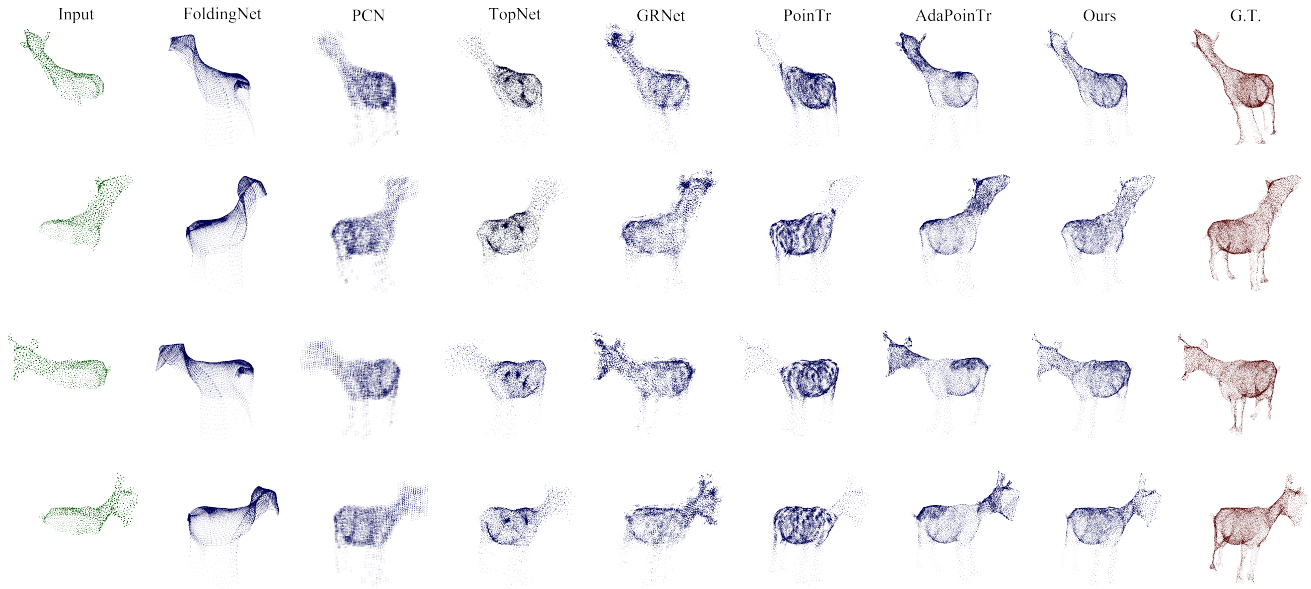


Fig. 8 Comparison with other point cloud completion methods in crop validation method

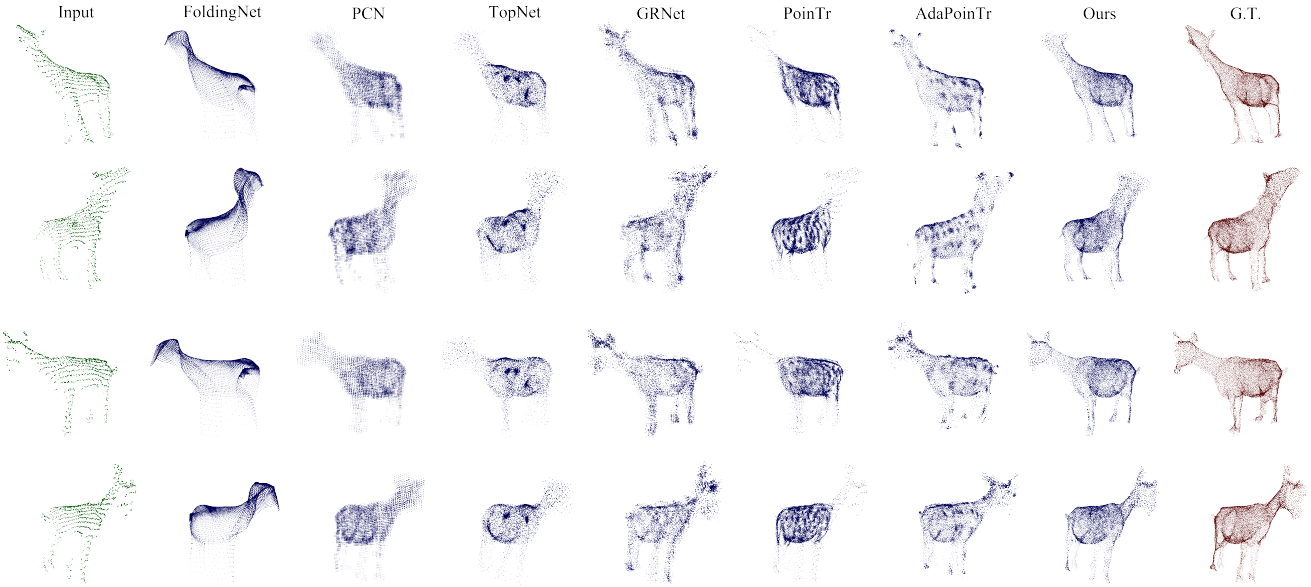


Fig. 9 Comparison with other point cloud completion methods in back-projecting verification approach

4.5 Ablation Study

In order to test the effectiveness of our design, we conducted a detailed ablation study on key components of the GoatPoinTr model. The results are summarized in Table 2. The baseline model A is the AdaPoinTr model used solely for point cloud completion. Replacing the traditional masking approach with our proposed back-projecting strategy (model B) showed a significant reduction in both versions of CD.

We found that replacing the self-attention mechanism with Outlook Attention [19] in AdaPoinTr encoder (Model C) improves all point cloud completion metrics, demonstrating its enhanced capability in capturing fine details and contextual relationships.

Table 2 Ablation study on the goat dataset (CD-l_1, CD-l_2 in units of 1e-3, MPJPE in units of 1e-2)

	F-Score@1% ↑	CD-l_1 ↓	CD-l_2 ↓	MPJPE ↓	PCK@0.05 ↑
A	0.555	10.768	0.396	-	-
B	0.602	10.105	0.309	-	-
C	0.644	9.544	0.271	-	-
D	0.637	9.590	0.276	2.325	0.9706
E	0.638	9.549	0.273	2.221	**0.9741**
Ours	**0.659**	**9.326**	**0.264**	**2.220**	0.9708

When we divided the global features for independent use in point cloud completion and keypoint detection tasks (model D), we observed a slight dip in performance metrics. This outcome suggests that the lack of interaction between these two tasks, post-feature separation, could be detrimental to the overall model efficiency. Subsequently, we integrated the coarse feature obtained from the coarse point through a learnable MLP into the keypoint feature (model E). The results showed significant improvements in both point cloud completion and keypoint detection tasks compared to model D. This shows that the coarse points do work for the keypoint detection task and that the results of the keypoint detection also work for the coarse point in back propagation, which leads to a better completion of the point cloud. Finally, the quality of point cloud completion is significantly improved by integrating the detected keypoints into the coarse point (ours), which also confirms the positive effect of the keypoint detection task on the point cloud completion.

4.6 Goat Body Measurement Results

For manual measurements, an experienced herdsman kept each goat calm and standing upright in an open area. A measuring stick was used for BL, BH, CW, CD, HH, and HW, and a tape measure for CG. Each dimension was measured five times per goat and averaged to reduce measurement noise.

Table 3 Comparison of MAE and MAPE for Goat Body Measurements

	single-view (ours)		multi-view [10]	
	MAE (cm) ↓	MAPE (%) ↓	MAE (cm) ↓	MAPE (%) ↓
BL	2.40	3.01	2.57	3.24
BH	2.76	3.83	1.83	2.54
CW	1.06	4.80	1.13	5.43
CG	5.07	5.55	2.85	3.08
CD	1.38	4.24	-	-
HH	1.99	2.66	1.62	2.16
HW	1.06	4.58	1.02	4.59

Table 3 compares the MAE and MAPE for body measurements between our single-view method and the multi-view method from [10], both evaluated on the same dataset. Despite using a single-view approach, our method achieves comparable accuracy, particularly for body length (BL), where length reference information is directly observable from the side view. All measurements, except CG, show MAPE values under 5%, demonstrating effective accuracy.

The CG measurement exhibited higher errors, suggesting challenges in fully recovering detailed chest information from a single view, which impacts the accuracy of chest girth estimation. This difficulty is further influenced by posture variations, underscoring the need for further refinement in chest dimension estimation.

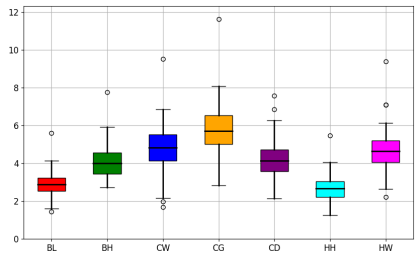


Fig. 10 MAPE boxplot of goats body measurements

Figure 10 shows the MAPE distribution of goat body measurements. Chest width (CW) and hip width (HW) have more outliers, mainly due to the lower bound of absolute error caused by camera calibration and keypoint detection precision. This effect is more noticeable in measurements with smaller actual distances, like CW and HW, leading to greater variations in MAPE.

5 CONCLUSIONS

In this paper, we proposed GoatPoinTr, a joint point cloud completion and keypoint detection network for automatic goat body size measurement from single-view point clouds. Based on the AdaPoinTr framework, we designed a back-projecting masking strategy to better simulate real-world single-view scenarios, introduced Outlook Attention to enhance fine-grained detail recovery, and developed a cross-optimization mechanism to enable mutual refinement between completion and keypoint detection tasks. We also established a goat point cloud benchmark with 5,712 samples and 10 annotated keypoints. Using the detected keypoints and completed point cloud, seven body dimensions were successfully measured, including body length , body height, chest width, chest girth, chest depth, hip height, and hip width. Experiments validated that GoatPoinTr outperforms existing completion methods and achieves body measurement accuracy comparable to multi-view approaches, with MAPE values of 2.66%–5.55%, demonstrating the feasibility of practical single-view livestock body measurement.

In precision breeding, our approach enables complete goat point cloud capture using handheld devices in non-channel environments, addressing multi-view collection limitations.Future work includes expanding this method to broader agricultural applications and exploring animal welfare and health monitoring via transfer learning, aiming to enhance production efficiency and sustainability.

Acknowledgements This work was supported by the National Key Research and Development Program of China (2022YFD1300200), the Xi'an Science and Technology Bureau Technology Tackling Special Plan Project (K4050724281),the Construction of "Scientist and Engineer" team of Qin Chuang Yuan in Shaanxi Province of China(2023KXJ-109), and the Doctoral Graduate Student Independent Innovation Research Project of Northwest A&F University (2025KYCXZ21).All data collection procedures were conducted in strict accordance with ethical standards, with no invasive interventions applied to the animals and no stress responses elicited.

Author Contributions B. J. and G. W. conceived and designed the overall methodology, implemented the proposed approach, and prepared the initial manuscript draft. C. H. was responsible for data acquisition and organization, and contributed to manuscript writing. J. F. participated in partial code development and optimization. Y. Q. reviewed the manuscript and contributed to manuscript writing. F. N. reviewed the manuscript and provided revision suggestions. M. W. supervised the overall project, contributed to manuscript writing and review. All authors reviewed and approved the final manuscript.

Data Availability The code and dataset used in this study are available upon reasonable request to the corresponding author.

Declarations

Conflict of 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.

References

1. Andrea Pezzuolo, Marcella Guarino, Luigi Sartori, Luciano A González, and Francesco Marinello. On-barn pig weight estimation based on body measurements by a kinect v1 depth camera. *Computers and electronics in agriculture*, 148:29–36, 2018.
2. Wentao Yuan, Tejas Khot, David Held, Christoph Mertz, and Martial Hebert. Pcn: Point completion network. In *2018 international conference on 3D vision (3DV)*, pages 728–737. IEEE, 2018.
3. Jianlong Zhang, Yanrong Zhuang, Hengyi Ji, and Guanghui Teng. Pig weight and body size estimation using a multiple output regression convolutional neural network: A fast and fully automatic method. *Sensors*, 21(9):3218, 2021.
4. Suvarna Bhoj, Ayon Tarafdar, Anuj Chauhan, Mukesh Singh, and Gyanendra Kumar Gaur. Image processing strategies for pig liveweight measurement: Updates and challenges. *Computers and Electronics in Agriculture*, 193:106693, 2022.
5. Jennifer Salau, Jan H Haas, Wolfgang Junge, and Georg Thaller. A multi-kinect cow scanning system: Calculating linear traits from manually marked recordings of holstein-friesian dairy cows. *Biosystems Engineering*, 157:92–98, 2017.
6. Alexey Ruchay, Vitaly Kober, Konstantin Dorofeev, Vladimir Kolpakov, and Sergei Miroshnikov. Accurate body measurement of live cattle using three depth cameras and non-rigid 3-d shape recovery. *Computers and Electronics in Agriculture*, 179:105821, 2020.
7. Hu Hao, Yu Jincheng, Yin Ling, Cai Gengyuan, Zhang Sumin, and Zhang Huan. An improved pointnet++ point cloud segmentation model applied to automatic measurement method of pig body size. *Computers and Electronics in Agriculture*, 205:107560, 2023.
8. Zixia Hou, Lyuwen Huang, Qi Zhang, and Yuanshuang Miao. Body weight estimation of beef cattle with 3d deep learning model: Pointnet++. *Computers and Electronics in Agriculture*, 213:108184, 2023.
9. Ke Wang, Hao Guo, Qin Ma, Wei Su, Luochao Chen, and Dehai Zhu. A portable and automatic xtion-based measurement system for pig body size. *Computers and electronics in agriculture*, 148:291–298, 2018.

10. Bo Jin, Guorui Wang, Jingze Feng, Yongliang Qiao, Zhifeng Yao, Mei Li, and Meili Wang. Pointstack based 3d automatic body measurement for goat phenotypic information acquisition. *Biosystems Engineering*, 248:32–46, 2024.
11. Lvwen Huang, Shuqin Li, Anqi Zhu, Xinyun Fan, Chenyang Zhang, and Hongyan Wang. Non-contact body measurement for qinchuan cattle with lidar sensor. *Sensors*, 18(9):3014, 2018.
12. XY Zhang, Gang Liu, Ling Jing, YS Si, XH Ren, and L Ma. Automatic extraction method of cow's back body measuring point based on simplification point cloud. *Transactions of the Chinese Society for Agricultural Machinery*, 50:267–275, 2019.
13. Ke Wang, Hao Guo, Qin Ma, W Su, MR Hua, and DH Zhu. Pig body size measurement method based on mirror reflection of single angle point cloud [j]. *Transactions of the Chinese Society for Agricultural Machinery*, 49(3):187–195, 2018.
14. Qi Zhang, Zixia Hou, Lyuwen Huang, Feiyi Wang, and Hongyan Meng. Reparation with moving least squares sampling and extraction of body sizes of beef cattle from unilateral point clouds. *Computers and Electronics in Agriculture*, 224:109208, 2024.
15. Peng Xiang, Xin Wen, Yu-Shen Liu, Yan-Pei Cao, Pengfei Wan, Wen Zheng, and Zhizhong Han. Snowflakenet: Point cloud completion with snowflake point deconvolution with skip-transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 5499–5509, 2021.
16. Haoran Zhou, Yun Cao, Wenqing Chu, Junwei Zhu, Tong Lu, Ying Tai, and Chengjie Wang. Seedformer: Patch seeds based point cloud completion with upsample transformer. In *European Conference on Computer Vision (ECCV)*, pages 416–432. Springer, 2022.
17. Xumin Yu, Yongming Rao, Ziyi Wang, Zuyan Liu, Jiwen Lu, and Jie Zhou. Pointr: Diverse point cloud completion with geometry-aware transformers. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 12498–12507, 2021.
18. Xumin Yu, Yongming Rao, Ziyi Wang, Jiwen Lu, and Jie Zhou. AdaPoinTr: Diverse Point Cloud Completion With Adaptive Geometry-Aware Transformers . *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 45(12):14114–14130, December 2023.
19. Li Yuan, Qibin Hou, Zihang Jiang, Jiashi Feng, and Shuicheng Yan. Volo: Vision outlooker for visual recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(5):6575–6586, 2023.
20. Kiyoun Kwon and Duhwan Mun. Iterative offset-based method for reconstructing a mesh model from the point cloud of a pig. *Computers and Electronics in Agriculture*, 198:106996, 2022.
21. Kwan Pang Tsui, Kin Hong Wong, Changling Wang, Ho Chuen Kam, Hing Tuen Yau, and Ying Kin Yu. Calibration of multiple kinect depth sensors for full surface model reconstruction. In *First International Workshop on Pattern Recognition*, volume 10011, pages 305–310. SPIE, 2016.
22. Keneni W Tesema, Lyndon Hill, Mark W Jones, Muneeb I Ahmad, and Gary K Sherley Tam. Point cloud completion: A survey. *IEEE Transactions on Visualization and Computer Graphics*, 30(10):6880–6899, 2024.
23. Lyne P Tchapmi, Vineet Kosaraju, Hamid Rezatofighi, Ian Reid, and Silvio Savarese. Topnet: Structural point cloud decoder. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 383–392, 2019.
24. Haozhe Xie, Hongxun Yao, Shangchen Zhou, Jiageng Mao, Shengping Zhang, and Wenxiu Sun. Grnet: Gridding residual network for dense point cloud completion. In *European Conference on Computer Vision*, pages 365–381. Springer, 2020.
25. Shanshan Li, Pan Gao, Xiaoyang Tan, and Mingqiang Wei. Proxyformer: Proxy alignment assisted point cloud completion with missing part sensitive transformer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9466–9475, 2023.
26. Jun Wang, Ying Cui, Dongyan Guo, Junxia Li, Qingshan Liu, and Chunhua Shen. Pointattn: You only need attention for point cloud completion. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 5472–5480, 2024.
27. Jincen Jiang, Xuequan Lu, Lizhi Zhao, Richard Dazeley, and Meili Wang. Masked autoencoders in 3d point cloud representation learning. *IEEE Transactions on Multimedia*, 27:820–831, 2023.
28. Liang Pan, Xinyi Chen, Zhongang Cai, Junzhe Zhang, Haiyu Zhao, Shuai Yi, and Ziwei Liu. Variational relational point completion network. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8524–8533, 2021.
29. Zhengyou Zhang. A flexible new technique for camera calibration. *IEEE Transactions on pattern analysis and machine intelligence*, 22(11):1330–1334, 2000.
30. Martin A Fischler and Robert C Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981.
31. Martin Ester, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu, et al. A density-based algorithm for discovering clusters in large spatial databases with noise. In *kdd*, volume 96, pages 226–231, 1996.
32. Svante Wold, Kim Esbensen, and Paul Geladi. Principal component analysis. *Chemometrics and intelligent laboratory systems*, 2(1-3):37–52, 1987.
33. Charles Ruizhongtai Qi, Li Yi, Hao Su, and Leonidas J Guibas. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. *Advances in neural information processing systems*, 30, 2017.
34. Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32, 2019.
35. Catalin Ionescu, Dragos Papava, Vlad Olaru, and Cristian Sminchisescu. Human3. 6m: Large scale datasets and predictive methods for 3d human sensing in natural environments. *IEEE transactions on pattern analysis and machine intelligence*, 36(7):1325–1339, 2013.
36. Yaoqing Yang, Chen Feng, Yiru Shen, and Dong Tian. Foldingnet: Point cloud auto-encoder via deep grid deformation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 206–215, 2018.