POINTNET++LR3D: AN IMPROVED POINTNET++ MODEL FOR INDIVIDUAL IDENTIFICATION OF PIG BACK POINT CLOUD
/
POINTNET++LR3D:一种改进的POINTNET++模型用于猪背点云个体识别

Yongshuai YANG[1)], Yaqi YAN[1)], Yuhang LI[1)], Jiarui ZHANG[1)], Xiaochan GAO[1)], Jie HU[1)], Juan LIU*[2)]
[1)] College of Software, Shanxi Agricultural University, Taigu, Shanxi / China;
[2)] Department of Basic Sciences, Shanxi Agricultural University, Taigu, Shanxi / China
Corresponding authors: Juan LIU; Tel: +86-18335463280; E-mail: liujuannk@sxau.edu.cn
DOI: https://doi.org/10.35633/inmateh-75-98

Keywords: *Individual identification; 3D point cloud; PointNet++; local context fusion; bilinear regularization*

ABSTRACT

Individual pig identification is a key technology to realize fine farming management, which is of great value in the fields of animal behavior tracking and health monitoring. Aiming at the limitations of traditional 2D vision methods in stereo feature extraction, this study uses pig back point cloud to effectively capture deep 3D features such as back contour and skin texture and proposes an improved PointNet++ model for pig individual identification, which explicitly captures the local geometric and feature differences through two-stream differential coding, refines the feature distribution by using low-rank bilinear decomposition and residual sharpening strategies, and then establish the two-way dependency between channel and space to generate the global perceptual map, and finally combine with Mish activation function to enhance the nonlinear feature extraction. The experiment takes the hybrid long white pig as the research object and uses the Intel D435i depth camera to collect data and construct the segmentation and identification model. The results show that the improved model PointNet++LR3D achieves an overall accuracy of 97.11% in the individual identification task, which is an improvement of 1.9% compared to the base PointNet++MSG model. In addition, extended tests on the ModelNet40 dataset show an improvement in classification accuracy to 93.1%, validating the generalization ability of the architectural improvements. This study provides an efficient solution for non-contact pig identification based on point cloud, demonstrating the potential for application in fine-tuned farming.

摘要

猪只个体识别是实现精细化养殖管理的关键技术，在动物行为追踪、健康监测等领域具有重要价值。针对传统二维视觉方法在立体特征提取上的局限性，本研究使用猪背点云有效捕捉背部轮廓和皮肤纹理等深度三维特征，提出了一种改进的PointNet++模型用于猪只个体识别，通过双流差分编码显式捕捉局部几何与特征差异，利用低秩双线性分解和残差锐化策略细化特征分布，进而建立通道与空间的双向依赖关系生成全局感知图，最后结合 Mish 激活函数增强非线性特征提取。实验以杂交长白猪为研究对象，使用 Intel D435i 深度相机采集数据，构建分割与识别模型。结果表明，改进模型 PointNet++LR3D 在个体识别任务中整体准确率达 97.11%，相比基础 PointNet++MSG 模型提升 1.9%。此外，在 ModelNet40 数据集上的扩展测试显示分类准确率提升至 93.1%，验证了架构改进的泛化能力。本研究为基于点云的非接触式猪只识别提供了高效解决方案，展现了在精细化养殖中的应用潜力。

INTRODUCTION

In modern agriculture and animal management, individualized management and fine monitoring have become the key factors in improving breeding efficiency and animal welfare (*Neethirajan et al., 2024*). Individual identification is the key technology to realize individual tracking of health data (*Vidal et al., 2021*), which provides the necessary support for various aspects of precision management, health monitoring, fine feeding, behavioral monitoring and analysis, breeding selection, and automated management (*Su et al., 2024; Krampe et al., 2024*).

The traditional animal individual identification technology mainly has two major categories: physical tag identification and radio frequency identification (*Ruiz et al., 2011*); currently applied in the breeding industry is still the most used in the physical tag using ear tags or collars to distinguish each body, however, the physical tags are prone to wear and tear, loss and damage, but also requires manual operation close contact with the animal, which increases the human cost as well as the risk of animal stress (*Martínez et al., 2016*).

With the rapid development of computer vision and deep learning technologies (*Wang et al., 2022*), vision-based individual identification methods have gradually become a research hotspot, which significantly improves the flexibility and robustness of identification mainly by capturing the external features of pigs, such as their faces, body sizes, or back contours (*Saleem et al., 2021; García et al., 2020*). However, 2D images are still unresolved due to factors such as light, angle and occlusion, which make it difficult to fully capture the three-dimensional morphological features of pigs, especially in dynamic scenes, and the stability of the identification effect still needs to be improved. In contrast, 3D point cloud technology (*Guo et al., 2020*) obtains three-dimensional information about pigs through depth sensors (*Yu et al., 2024*), which can more comprehensively describe the dorsal contour, curvature, and spatial structural features of pigs (*Shuai et al., 2020*), and has become a hot spot of research by its advantage of being able to obtain three-dimensional depth information of pigs.

In recent years, significant progress has been made in applying point cloud technology in livestock individual identification. Zhou and others first proposed an individual identification method based on the pig back point cloud, utilizing PointNet++ for back point cloud segmentation and constructing a pig individual identification model based on the improved PointNet++LGG algorithm by increasing the adaptive global sampling radius, deepening the network structure, and increasing the number of features, with accuracy reaching 95% (*Zhou et al., 2023*). Similarly, Kyaw et al. used the PointNet++ algorithm to detect and segment the dorsal surface region of a cow from a point cloud image, providing applications for individual cow identification, lameness detection, and body condition scoring (*Kyaw et al., 2024*). These studies achieved non-contact and high-precision individual identification through 3D point cloud technology, overcoming traditional methods' limitations and providing technical support for fine farming.

Inspired by the above studies, this study notes that the application of point cloud technology in livestock individual identification is still in its infancy and that most of the current improvements of PointNet++ focus on the feature extraction stage, especially on enhancing the expression of local features by optimizing the grouping of points such as multi-scale grouping or dynamic grouping (*Luo et al., 2024; Nong et al., 2023*). However, fewer studies have systematically improved the characterization ability of point cloud features by starting from the fusion mechanism of local and global features. Therefore, this study proposes a new improvement strategy based on this, constructing an innovative module of local context fusion and global bilinear regularization, integrating it into the network architecture of PointNet++, explicitly modeling geometric and feature differences within the framework of PointNet++, and refining the feature distributions through global sensing, and utilizing low-rank bilinear decomposition and residual sharpening strategies to refine the feature distributions. Feature distributions to establish the bidirectional dependence between channel and space. This approach not only enhances the model's sensitivity to the local details of the pig back point cloud, such as curvature and edges, but also improves the feature differentiation through global regularization, thus achieving higher accuracy and robustness in the task of livestock individual identification in dynamic scenes.

MATERIALS AND METHODS

Introduction of experimental pig data

The data on the pigs used in this experiment were collected from Huifeng Breeding Professional Cooperative in Guangling County, Datong City, Shanxi Province, and the experimental pigs were labeled and approved by the pig farm. The pig farm features a semi-enclosed structure, and the experimental data were collected under two natural conditions: with sunlight and without sunlight. In order to increase the robustness of the experiment, no external lights were used throughout the collection, and point cloud data were collected from the backs of ten pigs, all of which were crossbred long white pigs and pigs with similar age, weight, and body hair color were selected as the collection objects.

Data Acquisition

In this study, the following acquisition plan was developed during the acquisition process, as shown in Fig.1. The depth camera Intel D435i was used to collect the depth information of the pig's back, the D435i adopts a dual-camera design and calculates the depth information based on binocular stereo vision (*Servi et al., 2024*; *Zhang et al., 2022*), which generates high-resolution depth data by capturing the parallax images from both cameras. In order to capture individual pigs in multiple postures, the recorded video was used to capture depth information and RGB color images simultaneously from the upper part of the pig's back.

To verify whether different camera heights and angles affect identification accuracy, the camera positions were intentionally varied. The cameras were mounted on the top of a telescopic pole, which was handheld to capture approximately one minute of video from each angle of the pig's back. The videos were recorded at a resolution of 468 × 828 pixels and a frame rate of 30 Hz. In total, 10 video recordings in the BAG format were collected.

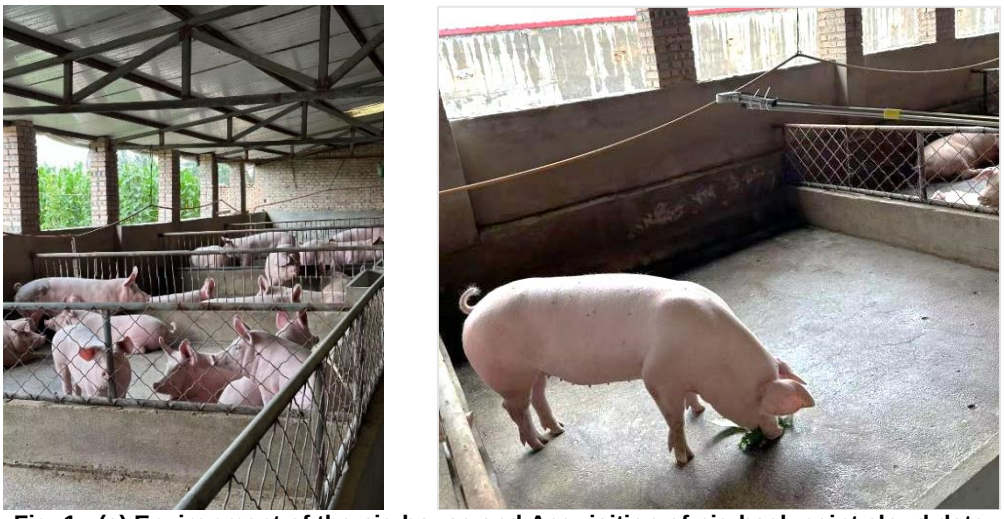


Fig. 1 - (a) Environment of the pig house and Acquisition of pig back point cloud data

Data pre-processing

(1) Depth image to point cloud image conversion

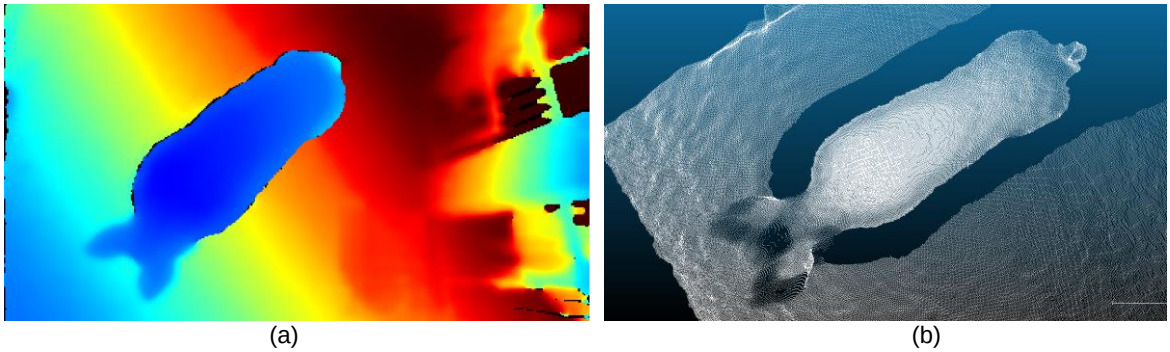
Each BAG file stores the depth and RGB color information of the ten pigs. Additionally, the camera's intrinsic parameters (focal lengths f_x, f_y and principal point coordinates C_x, C_y) are recorded within each file. These parameters are extracted by parsing the BAG files.

First, the BAG files are parsed frame by frame using Intel RealSense to extract depth and color frames. Depth images are manually filtered with the assistance of corresponding color images; only those with poor-quality back data are removed to ensure the reliability of the experimental dataset. Since the objective of this experiment is to assess whether individual pigs can be accurately identified based on body features - such as the contours and curvature of the back - rather than color features, only the depth images are retained after filtering. These retained depth images are subsequently converted into point cloud representations.

The depth value of each pixel in the depth image is converted to the actual 3D coordinates by combining the camera's internal reference, assuming that the pixel coordinates are U and V, and the actual depth is Z. The 3D coordinates (X, Y, Z) of the point can be expressed as:

$$X = \frac{(U-C_x)\cdot Z}{f_x}, \quad Y = \frac{(V-C_y)\cdot Z}{f_y}, \quad Z = depth\_image(U,V) \tag{1}$$

where: $(U - C_x)$ and $(V - C_y)$ denote the horizontal and vertical distances of the pixel point relative to the center of light, and the X Y coordinates of the pixel point in the actual 3D space can be obtained by multiplying the distance by the depth value of the point, Z, and dividing by the focal length, f_x or f_y.



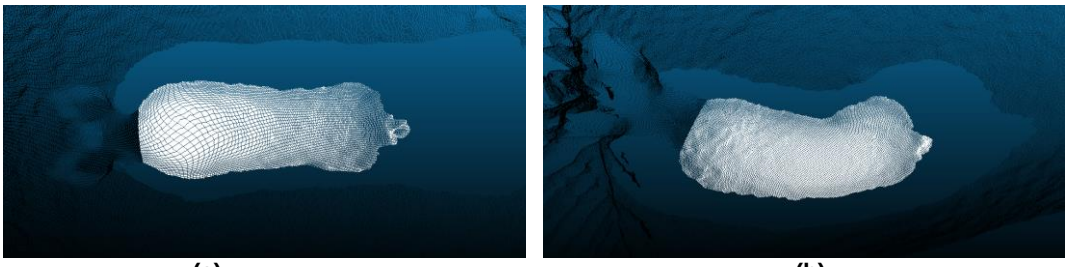
(a) (b)

Fig. 2 - Depth images converted to point cloud images

(a) Pig depth image (b) Pig point cloud image

(2) Pig back point cloud labeling

The generated dorsal point cloud data not only has the pig body point cloud but also includes a large number of background point clouds, using CloudCompare software to label the point clouds as pig dorsal point clouds and background point clouds because the pig is in a low head posture most of the time and the head bobbing is more obvious, most of the captured head point cloud images are incomplete and badly adhered to the ground and walls, in order to avoid the interference of the head point cloud, the head is labeled as the background point when labeling the point clouds, only the dorsal point cloud is retained as the segmentation target. As shown in Fig. 3, the pig has two main postures, standard standing and twisted body, and the two points with the most considerable curvature changes between the neck and the back of the pig are segregated for labeling.



(a) **(b)**

Fig. 3 - Pig back point cloud labelling in two poses

(a) Splitting of a pig in standard standing position(b) Segmentation of a pig with twisted body

Segmentation and identification Modelling

The process of pig individual identification based on pig back point cloud is mainly composed of two parts: segmentation model construction and identification model construction. First, the collected raw data are converted into point cloud images, then the background and head are removed by the segmentation model. Finally, the segmented pig back point cloud is input into the individual identification model, and the specific number of the pig is output.

Back - background point cloud segmentation

(1) Segmentation method

The PointNet family of networks are neural networks specialized in point cloud data processing that directly manipulate irregular point cloud data without the need to convert the point cloud data into regular grids or voxels (*Qi et al., 2017*). The segmentation problem in this study is relatively simple, with only two parts, the back and the background, but the data used in this study is a dynamic moving image of a pig, and the posture of the pig at the neck is severely deformed when it is moving. Compared with PointNet, PointNet++ extracts local features layer by layer through the process of 'sampling-grouping-feature extraction-feature propagation,' which is more delicate processing of local features, so the PointNet++ network is chosen to build the segmentation model in this study.

(2) Segmentation model based on PointNet++MSG

As shown in Fig. 4, the segmentation network chooses the PointNet++ Multiscale Grouping Feature Learning (PointNet++MSG) framework, which includes four parts: two-layer sampling, multiscale grouping, feature extraction downsampling and feature propagation upsampling (*Qi et al., 2017*).

Due to the vast number of points in the original point cloud, random sampling of the original point cloud is required, and the network receives the randomly sampled globally uniformly sampled points containing xyz coordinates and normal vectors. Feature extraction is performed by three Set Abstraction (SA) layers downsampling step by step: SA1 receives the 6-dimensional point cloud input, extracts 64-, 128-, and 128-dimensional features and splices them in [0.1,0.2,0.4] triple radius on 512 sampling points; SA2 generates 256-dimensional features in [0.4,0.8] double radius on 128 sampling points; SA3 generates 256-dimensional features by global pooling to generate 1024-dimensional global features. Feature propagation is performed by Feature Propagation (FP) with layer-by-layer upsampling: FP3 fuses global and SA2 features, FP2 reconstructs the mid-level geometry, and FP1 splices the original point coordinates with labels to recover fine-grained features. The classification head uses a two-layer convolutional network to output semantic category probability distributions for pig back and background.

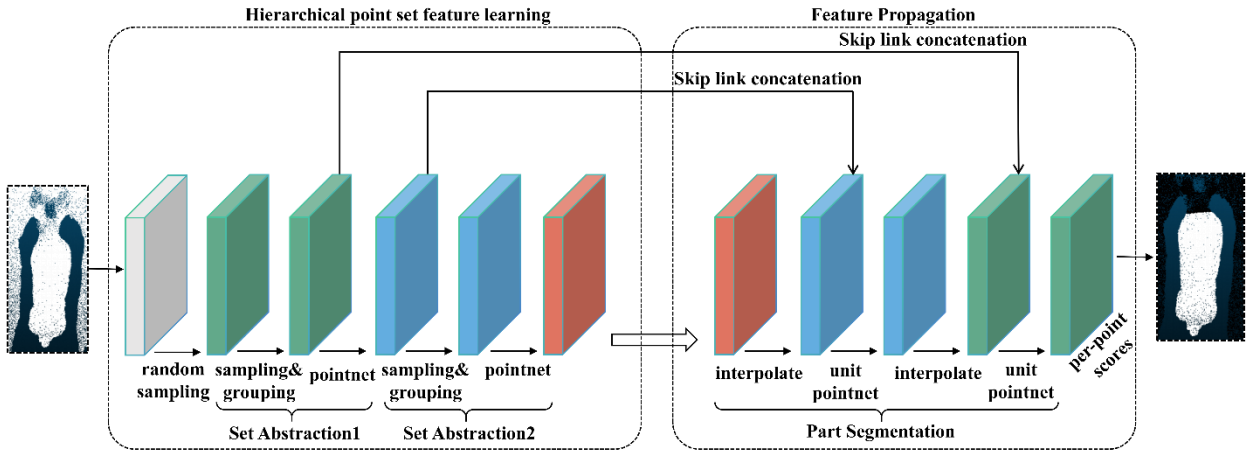


Fig. 4 - Segmentation model for pig back based on PointNet++MSG

Individual identification of pigs

(1) Individual identification method

Compared with segmentation, the identification task needs to focus on local and global feature extraction. In this study, individual identification models for pigs are developed based on the single-scale (SSG) and multi-scale (MSG) grouping strategy algorithms of PointNet and PointNet++, respectively, as the baseline model for individual identification. An improved model PointNet++LR3D with local context fusion as well as global bilinear regularization is proposed, which significantly improves the characterization of high-dimensional features of pig dorsal point clouds by means of dual-stream differential coding of geometries and features, low-rank bilinear decomposition and residual sharpening strategies.

(2) Individual pig identification model based on PointNet++MSG

As shown in Figure 5, the individual pig identification model and segmentation model are similar in principle, except that compared to the segmentation task, the individual identification only has 'sampling-grouping-feature extraction' without the process of feature propagation, and the 1024-dimensional global features obtained after three-layer SA feature extraction are directly input into the classification layer, and the classification layer outputs the category of each pig through a three-layer full connection. The classification layer outputs the class distribution probability of each pig.

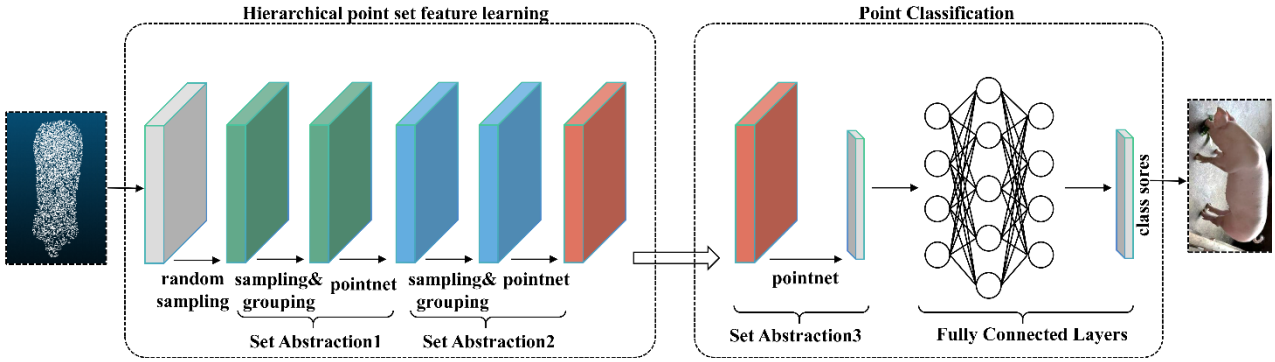


Fig. 5 - Individual identification model for pigs based on PointNet++MSG

Local Context Fusion with Bilinear Regularization Strategies for LR3D Modules

In traditional point cloud data processing networks, PointNet++ only aggregates MLP-encoded features in local coordinates, ignoring differences in features between neighboring points, and PointNet++'s Set Abstraction aggregates local features through farthest-point sampling and grouping and extracts features only through MLP stacking, does not explicitly model point-neighborhood relationships, and lacks the sensitive capture of geometric differences such as curvature and regular vector changes. EdgeConv, on the other hand, mainly performs dynamic graph convolution based on neighborhood point feature differences without directly modeling 3D geometric coordinate relationships. As shown in Figure. 6, this study uses a CNN-based model to learn feature maps, proposes dual-stream differential coding to model both geometric and feature differences to capture local structural changes explicitly, and enables geometric and feature contexts to be encoded separately through shared MLP and asymmetric pooling using split aggregation to avoid confusion between coordinate and feature information and enhance interpretability.

(1) Local context fusion strategy for two-stream feature splicing

To construct the local context for each point, the k-nearest neighbor (KNN) algorithm is employed to identify its neighborhood, which is a widely used neighborhood finding method in point cloud processing. The KNN algorithm calculates the 3D Euclidean distances from all points in the point cloud to a target point p_i and selects the k closest points as its neighbors: $\mathcal{N}(p_i) = \{p_{i1}, p_{i2}, \dots, p_{ik}\} \in N(p_i)$, by combining the edge $\forall p_{ik} \in N(p_i)$ between p_i itself and its k neighbors, the local geometric map in 3D space is: $\tilde{p} = [p_i; p_{ik} - p_i] \in \mathbb{R}^{k\times 6}$, where p_i is the current point, and $p_{ik} - p_i$ is the coordinate difference reflecting the local geometric structure. Thus, the local geometric map of all points P is denoted as $\tilde{P} = [p_1, p_2, \dots, p_i, \dots, p_N,] \in \mathbb{R}^{N\times k\times 6}$. Finally, the local geometric map is encoded by shared MLP, and the local geometric context encoding is aggregated by applying the Max Pooling function to k neighbors:

$$P = \max_k \left(MLP_\Theta(\tilde{P})\right), P \in \mathbb{R}^{N\times \frac{C}{2}} \quad (2)$$

Meanwhile, the local feature map of f_i in C-dimensional space can be formed: $\tilde{f} = [f_i; f_{ik} - f_i] \in \mathbb{R}^{k\times 2C}$, where f_{ik} is the corresponding feature of $p_{ik} \in N(p_i)$, and $f_{ik} - f_i$ is the feature difference, which reflects the semantic difference. Therefore, the local feature map of a feature map F is represented as $\tilde{\mathcal{F}} = [f_1, f_2, \dots, f_i, \dots, f_N] \in \mathbb{R}^{N\times k\times 2C}$. The local feature context encoding is obtained by following a similar operation in Equation (2):

$$\mathcal{F} = \max_k \left(MLP2_\Theta(\tilde{\mathcal{F}})\right), \mathcal{F} \in \mathbb{R}^{N\times \frac{C}{2}} \quad (3)$$

Where $MLP2_\Theta$ is another shared MLP encoding the local feature map, finally the local geometric context and the local feature context() are concatenated as the output of the final local context fusion block:

$$\mathcal{F}_L = concat(P, \mathcal{F}), \mathcal{F}_L \in \mathbb{R}^{N\times C} \quad (4)$$

Compared with EdgeConv operations, the method proposed in this study defines and constrains both local geometric context and feature context and has inherent 3D geometric relationships that can directly reflect local surface geometric changes and semantic differences. In addition, local context fusion blocks can be flexibly deployed at different point cloud resolutions and CNN layers, which benefits most existing point cloud networks.

(2) Global bilinear regularization

In addition to collecting more local details for the feature representation of each point, the global bilinear regularized blocks proposed in this study aim to refine the feature maps by considering the global perception of the entire point cloud. In conventional self-attention mechanisms, global perception is estimated as long-range dependencies between point features, i.e., cosine similarity, which consumes significant memory. In contrast, the global perception in this study is computed as element-wise dependencies between feature maps based on global channels and point descriptors, and through a low-rank decomposition strategy, the memory consumption is significantly reduced.

In order to encode the global channel descriptor, firstly, the input features are linearly transformed by applying the weight matrix $W_C \in \mathbb{R}^{C\times \frac{C}{r}}$, where r is the reduction factor to reduce the dimension of the fused output $\mathcal{F}_L \in \mathbb{R}^{N\times C}$; then, the ReLU function is utilized to not only provide the nonlinearity after linearly mapping with W_C but also to satisfy the requirement of non-negativity in Equation (7); lastly, through the average pooling operation on the N elements along the spatial axis, the spatial information can be compressed into the global channel descriptor g_c. The above operation is as follows:

$$g_c = \operatorname*{avg}_N\left(ReLU(\mathcal{F}_L W_C)\right), g_c \in \mathbb{R}^{\frac{C}{r}} \quad (5)$$

Where '$\mathcal{F}_L W_C$' is the matrix product between $\mathcal{F}_L$ and W_C, and 'avg' denotes the average pooling along the spatial dimension N with a reduction factor satisfying r≥2. Furthermore, $g_c = \left[\mu_1, \mu_2, \dots, \mu_j, \dots, \mu_{\frac{C}{r}}\right]$, where μ_j represents the global response of the j-th channel in the whole point cloud feature map.

The global point-by-point descriptor g_p can also be generated using a similar approach in the context of another weight matrix $W_p \in \mathbb{R}^{C\times \frac{C}{r}}$, the ReLU function, and an average pooling operation on the $\frac{C}{r}$ elements along the channel axis:

$$g_p = \operatorname*{avg}_{\frac{C}{r}}\left(ReLU(\mathcal{F}_L W_p)\right), g_p \in \mathbb{R}^N \quad (6)$$

In addition, $g_p = [\lambda_1, \lambda_2, \ldots, \lambda_i, \ldots, \lambda_N]$,where λ_j is the global response of the i-th point in the whole point cloud feature map.

Unlike the use of Hadamard product between vectors, the geometric mean $\sqrt{\lambda_i \mu_j}$ captures the higher-order interactions of the channel with the space, providing a more robust global response than arithmetic averaging, and captures the low-rank global bilinear response by taking the square root of the outer product of g_p and g_c:

$$G = sqrt(g_p \otimes g_c), G \in \mathbb{R}^{N \times \frac{C}{r}} \quad (7)$$

where the element η_{ij} located in the ith row and jth column of G is mathematically computed as:

$$\eta_{ij} = \sqrt{\lambda_i \mu_j}, \eta_{ij} \in \mathbb{R} \quad (8)$$

The synthesis of the global bilinear response of all elements according to the corresponding point-by-point and channel descriptions in Equation 6 can be interpreted as an effective and efficient method in the following way. For each element, λ_i and μ_j are the arithmetic mean of the ith point and jth channel, respectively, and η_{ij} is the geometric mean of λ_i and μ_j. Firstly, taking the square root enhances the numerical stability, and secondly, it captures the bi-directional dependence of the channel on the space and provides a higher-order averaged response based on the spatial and channel-related information.

(3) Residual Sharpening Feature Recovery

After restoring the channel dimensions using a shared MLP, a residual connection is employed to preserve local features. Finally, the channel dimensions are fully recovered, and a full-size global perceptual map is generated:

$$\mathcal{F}_G = MLP(G + \mathcal{F}_L W_C + \mathcal{F}_L W_p), \mathcal{F}_G \in \mathbb{R}^{N \times C} \quad (9)$$

The average pooling operations in Equations 4 and 5 are used to compress global information from the point cloud and channel space, respectively; however, average pooling, as a conventional method for feature map compression, tends to generate smooth global representations that may weaken the uniqueness and representativeness of the features. To this end, this study proposes a more effective global-aware feature exploitation strategy to enhance feature differentiation in point cloud analysis. Specifically, to sharpen the learned features, the uniqueness of local features is emphasized by subtracting the global perceptual feature $\mathcal{F}_G$ from the local context fusion output $\mathcal{F}_L$. This subtraction effectively filters out redundant patterns originating from the global average vectors g_p and g_c, thereby reinforcing the distinctiveness of local features. The use of the Mish activation function further enriches the final output feature map by introducing additional nonlinear transformations.

$$\mathcal{F}_{out} = Mish(\mathcal{F}_L - \mathcal{F}_G), \mathcal{F}_{out} \in \mathbb{R}^{N \times C} \quad (10)$$

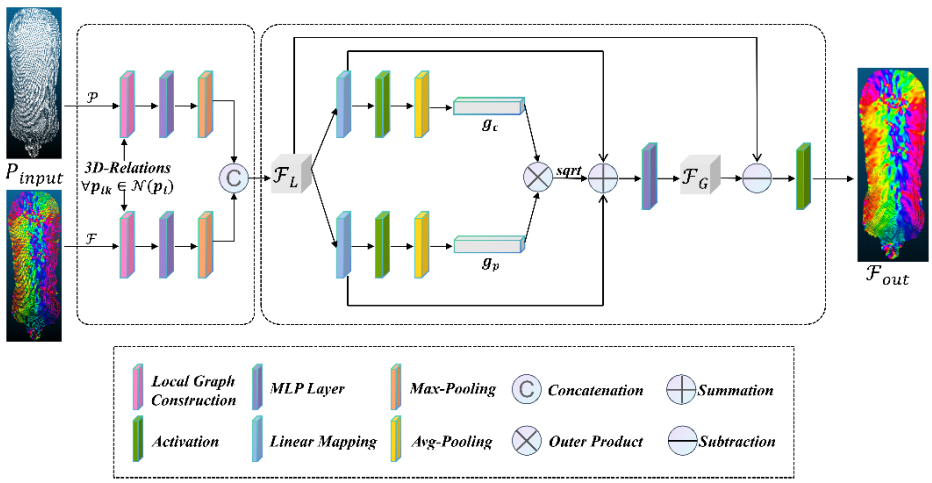


Fig. 6 - LR3D module introducing local context fusion and bilinear regularization

Integration of LR3D module in PointNet++MSG network

The LR3D (Local-Regional 3D) module proposed in this study is a novel point cloud feature extraction mechanism that enhances point cloud feature representation through two key components: local context fusion and global bilinear regularization. The LR3D module is integrated into the PointNet++(MSG) network architecture, and feature enhancement is mainly performed at two key layers, as shown in Figure 7, in the Set

Abstraction (SA) layer of PointNet++(MSG), by inserting the LR3D module after the SA1 and SA2 layers, respectively, with the following improvement strategies:

The LR3D module is introduced after the SA1 layer to explicitly capture local geometric and feature differences through a two-stream differential coding strategy for local context fusion. Local context encoding $\mathcal{F}_L$ is generated using a shared MLP with asymmetric pooling, while the global perceptual map $\mathcal{F}_G$ is produced via a global bilinear regularization block applied to the dual feature streams. This process refines the feature distribution using low-rank bilinear decomposition and a residual sharpening strategy, establishing a bidirectional dependency between channel and spatial dimensions. As a result, the network gains a stronger understanding of the relationship between global structures and local details in the pig back point cloud. The final output feature map $\mathcal{F}_{OUT}$ undergoes nonlinear transformation through the Mish activation function, further enhancing the model's robustness and generalization capability in complex point cloud scenarios, such as distortions caused by pig movement and variations in height. This approach strengthens the SA1 layer's ability to capture low-level local structures (e.g., back edges and body contours), addressing the limitation of traditional PointNet++ architectures that rely solely on stacked MLPs, and improves both global consistency and individual differentiation. The LR3D module is subsequently applied after the SA2 layer to further integrate the high-dimensional features extracted by SA1. Finally, the feature stream enhanced by the two LR3D modules is passed through a third Set Abstraction (SA3) layer for global feature aggregation, followed by classification via a multilayer perceptron (MLP).

This staged integration approach makes full use of the multi-scale grouping properties of the SA1 and SA2 layers, enabling the LR3D module to optimize the local and global representation capabilities in the feature extraction stages at different resolutions, respectively. The experimental results show that after adding the LR3D module after the SA1 and SA2 layers, the model's accuracy in the individual pig identification task is significantly improved while maintaining the balance of computational efficiency, which fully verifies the effectiveness of this improvement strategy.

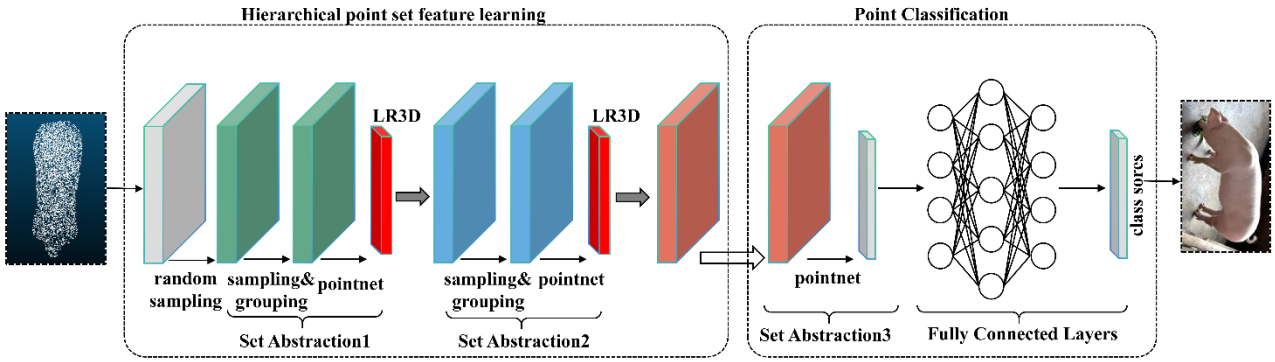


Fig. 7 - PointNet++ MSG classification model incorporating LR3D module

Experimental Sample Setting

This experiment collected 8,300 point cloud images from 10 bags and processed them to obtain the experimental dataset for 10 pigs. The dataset was divided into training, validation, and test sets in a ratio of 7:1.5:1.5. The training set was used for model learning, the validation set was used for parameter adjustment and performance evaluation, and the test set was used to verify the model's generalization ability.

Experimental parameter settings

The operating system used in this experiment is Ubuntu 20.04, the CPU is Intel Xeon Platinum 8270 2.70GHz, the memory is 64GB, the GPU is NVIDIA GeForce RTX 4090 D, the video memory is 24GB, and the deep learning frameworks are PyTorch 2.4.1 and Python 3.8.

Considering the impact of segmentation and identification speed, the initial sampling and target point numbers for the segmentation model are set to 1024 and 5000, respectively, while the sampling point number for the identification model is set to 2048. The batch size (Batch Size) is set to 8 for both models. The total number of training iterations (Epochs) for the segmentation and identification models is set to 100 and 200 cycles, respectively. The initial learning rate is set to 0.001, using the Adam optimizer with a weight decay rate of 1e-4. The learning rate is adjusted via a StepLR scheduler, decreasing to 0.7 of the original value every 20 cycles. The loss function employs negative log-likelihood loss (NLL Loss). The experiment is run on a single GPU, with data loading using 10 worker threads to enhance efficiency.

RESULTS

Experimental evaluation indexes

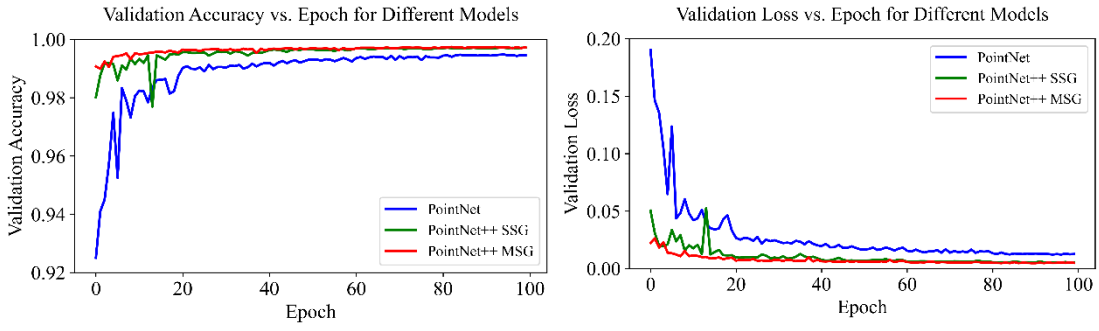
In the segmentation test, six metrics, namely, overall accuracy (OA), mean intersection and merger ratio (mIoU), precision, recall, F1 score, and average segmentation time (AvgTime) for pig back categories, were used to evaluate the segmentation performance of the model for pig back and background. In the identification (classification) test, the six metrics of OA, Precision, Recall, F1, Top-3 Accuracy, and Average identification Time (AvgTime) were used to evaluate the model's identification performance for individual pigs.

In the segmentation test calculation process, four key metrics are primarily used: true positive (TP), false positive (FP), true negative (TN), and false negative (FN). True positive (TP) represents the number of points predicted as pig back that are actually pig back. False positive (FP) represents the number of points predicted as pig back that are actually background. True negative (TN) represents the number of points predicted as background that are actually pig back. False Negative (FN) represents the number of points predicted as background but actually being pig back. In the identification test, TP represents the number of pigs correctly identified as belonging to a specific pig category, FP represents the number of pigs incorrectly identified as belonging to that category, TN represents the number of pigs correctly identified as belonging to the category but not actually belonging to it, and FN represents the number of completely incorrect identifications.

$$Over\ Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$
$$IOU = \frac{\mathrm{TP}}{TP + FP + FN}$$
$$Precision = \frac{TP}{TP + FP}$$
$$Recall = \frac{TP}{TP + FN}$$
$$F1 - score = \frac{2(Precision * Recall)}{Precision + Recall} \quad (11)$$

Pig back-Background Segmentation Model training results

In the pig back-background segmentation experiment, this study trained three models on the training and validation sets. Each model's validation accuracy and loss changed during training, as shown in Figure 8. All models' accuracy and loss values steadily increased and decreased with the increase in training iterations and gradually stabilized and converged after approximately 40 iterations. Among them, the PointNet++ MSG and PointNet++ SSG models performed well, with validation accuracy reaching approximately 99.8% and loss values tending towards 0, while the PointNet model performed slightly worse. Therefore, no significant overfitting was observed during training in the models. Overall, PointNet++ MSG achieves the best segmentation performance under the multi-scale grouping strategy, effectively capturing the local geometric features of pig back point clouds, thereby demonstrating stronger robustness and stability in segmentation tasks.



(a) **(b)**

Fig. 8 - Segmentation model validation results

(a) Validation accuracy curves(b) Validation loss curve

Pig back-Background Segmentation model test results

To further explore the segmentation performance of the models on unfamiliar datasets, the segmentation performance of the three models was evaluated using a test set, and the results are shown in Table 1.

PointNet++ MSG achieved 99.83% in Overall Accuracy (OA), which is an improvement of 0.03% to 0.30% compared to the other models. In terms of mean intersection and merger ratio (mIoU), PointNet++ MSG achieved 99.35%, which is 0.10% to 1.05% higher than other models, showing stronger segmentation consistency. In terms of Precision, Recall, and F1 scores, PointNet++ MSG achieved 99.38%, 99.53%, and 99.46%, respectively, which were 0.15% to 1.25%, 0.03% to 0.55% and 0.10% to 0.90% higher than the other models, indicating that it has a significant advantage in reducing misclassification and missed classification with significant advantages. Considering the real-time task requirements of individual identification, in terms of segmentation time (AvgTime), PointNet has the fastest inference speed of 0.004 s/sample but performs relatively poorly regarding segmentation accuracy. Although the inference time of PointNet++ MSG is slightly higher than the other two models at 0.071 s/sample, it still meets the real-time requirements.

In summary, PointNet++ MSG shows excellent performance in the pig back-background segmentation task. This task effectively deals with the complex geometric changes in the dynamic point cloud data and provides a high-quality segmentation basis for the subsequent individual identification task.

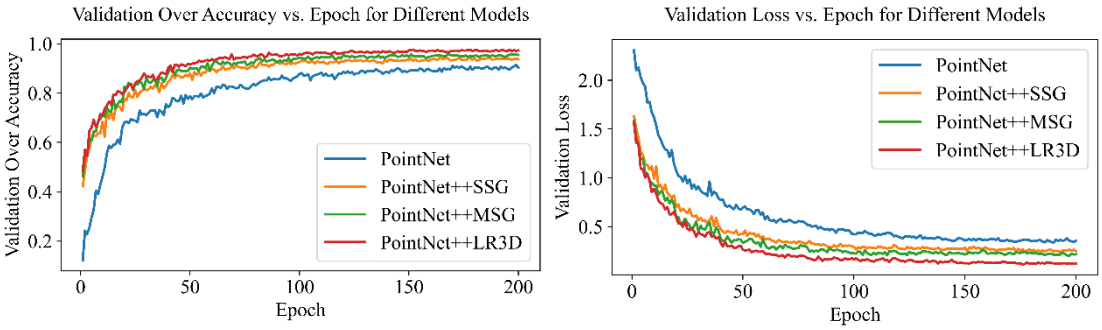
Table 1

Segmentation model test results

Model	OA [%]	mIoU [%]	Precision [%]	Recall [%]	F1 [%]	AvgTime [s/pig]
PointNet	99.53	98.30	98.13	98.98	98.56	0.004
PointNet++SSG	99.80	99.25	99.23	99.50	99.36	0.064
PointNet++MSG	99.83	99.35	99.38	99.53	99.46	0.071

Individual Pig Identification Model training results

In order to evaluate the performance of different models in the pig back point cloud individual identification task, the validation loss and accuracy of the base models (PointNet, PointNet++SSG, PointNet++MSG) and the improved PointNet++LR3D model were compared during the training process. The trend of the validation accuracy and loss of each model is shown in Figure 9. The experimental results show that the loss of all models decreases rapidly at the beginning of training and then levels off. At the same time, the accuracy increases rapidly and stabilizes at the end of training. The base PointNet model has the worst performance with higher test loss and lower accuracy; PointNet++SSG and PointNet++MSG have reduced loss and improved accuracy through hierarchical feature extraction. The improved PointNet++LR3D model performs best with the lowest test loss and highest accuracy. This excellent performance is attributed to the LR3D module's significant improvement in feature extraction through bilinear regularization of geometry and features, which better captures the complex geometric properties of pig dorsal point cloud data and provides a more efficient solution for individual identification tasks.



(a) **(b)**

Fig. 9 - Individual Identification model validation results

(a) Validation accuracy curve (b) Validation loss curve

Individual Pig Identification model testing results

In order to further verify the performance of different models in the pig dorsal point cloud individual identification task, the optimal training models were selected to test the PointNet, PointNet++SSG, PointNet++MSG, and the improved PointNet++LR3D models.

The overall performance indexes of each model in the individual identification task and the accuracy of individual identification for 10 pigs are shown in Tables 2 and 3.

The experimental results show that the basic PointNet model has an overall accuracy of 88.92% and an average identification time of 0.0015 s, which exhibits very high computational efficiency. However, its accuracy ranges from 75.89% to 95.24% from Pig1 to Pig10, with large fluctuations, and some individuals are poorly identified, especially Pig7 (75.89%), reflecting PointNet's limited ability to extract pig dorsal point cloud features. PointNet++SSG and PointNet++MSG through hierarchical feature extraction significantly improved the performance, where the overall accuracy of PointNet++SSG was 92.93% and that of PointNet++MSG was further improved to 95.42%, and the individual identification accuracies of the two were more balanced, ranging from 88.46% to 96.67% and 92.73% to 99.17%, respectively, and the individual identification accuracies of both were more balanced, ranging from 88.46% to 96.67% and 92.73% to 99.17%, respectively, due to the hierarchical feature extraction layer is introduced, the identification time grows to 0.0529 s and 0.0540 s. The improved PointNet++LR3D model proposed in this study performs the best, with the highest overall accuracy of 97.11%, a precision rate of 97.04%, a recall rate of 97.09%, an F1 score of 97.06%, and a Top-3 accuracy rate of 99.44%, the best performance in all indicators, and its accuracy on 10 pigs ranges from 95.38% to 99.21%, with the slightest fluctuation and the most stable identification effect. Due to the inclusion of the two-layer LR3D module, although the average inference time is slightly elevated to 0.0576 seconds, it still meets the real-time time requirement of the individual identification task. The LR3D module proposed in this study significantly enhances the model's ability to fine-grain feature extraction from point cloud data through geometric and feature bilinear regularization. Although the inference time of PointNet++LR3D is slightly higher than that of the other models, it strikes a good balance between performance enhancement and computational efficiency. It provides a better solution for the individual identification task in pigs.

Table 2

Overall test results of the different models under the individual identification task

Model	OA [%]	Precision [%]	Recall [%]	F1 [%]	Top-3 [%]	AvgTime [s/pig]
Pointnet	88.92	89.04	88.82	88.74	97.51	0.0015
PointNet++SSG	92.93	92.97	92.96	92.88	98.31	0.0529
PointNet++MSG	95.42	95.49	95.42	95.42	98.47	0.0540
PointNet++LR3D	97.11	97.04	97.09	97.06	99.44	0.0576

Table 3

Individual identification accuracy of each pig under different models

Model	Accuracy									
	[%]	[%]	[%]	[%]	[%]	[%]	[%]	[%]	[%]	[%]
pig	Pig1	Pig2	Pig3	Pig4	Pig5	Pig6	Pig7	Pig8	Pig9	Pig10
Pointnet	95.24	86.81	87.69	90.15	94.12	93.20	75.89	87.30	83.64	94.17
PointNet++SSG	93.33	93.06	88.46	91.67	92.44	94.56	90.18	92.86	96.36	96.67
PointNet++MSG	95.24	94.44	94.62	94.70	95.80	95.92	95.54	96.03	92.73	99.17
PointNet++LR3D	97.14	95.83	95.38	97.73	98.32	97.96	95.54	99.21	95.45	98.33

Confusion Matrix

In order to more intuitively assess the performance of the different models in the pig dorsal point cloud individual identification task, a classification confusion matrix was constructed for the four models, as shown in Figure 10. The horizontal axis of the confusion matrix represents the predicted labels, the vertical axis represents the actual labels, and the values on the diagonal indicate the number of correctly predicted samples. In contrast, the off-diagonal values indicate the number of misclassified samples. The results show that all models have high identification performance on most individuals, with correctly predicted samples mainly concentrated on the diagonal and fewer misclassified samples, verifying the effectiveness of the pig dorsal point cloud in the individual identification task.

By comparing the confusion matrices of the four models, it can be found that the base model has significant confusion in some categories. Among them, PointNet performed relatively poorly, with multiple misclassifications to other classes for pig2, 8, 9, and 6 misclassifications to pig4 for the predictions of pig3, pig7, and pig8, and bidirectional confusion between pig5 and pig9, with the error numbers of 7 and 4, respectively.

PointNet++SSG was significantly optimized by a single-scale grouping of the categorical purity of pig3, pig7, and pig8 and eliminated the one-way confusion of misclassification of pig9 to pig5. However, there were still four misclassifications of pig5 to pig9. PointNet++MSG further reduced the confusion of some categories by introducing multiscale grouping, significantly reduced the occurrence of misclassification of pig2, and further eliminated the two-way confusion between pig5 and pig9 bi-directional confusions.

In contrast, the improved model PointNet++LR3D shows significant advantages, with an average improvement of 10.7% in the number of correct diagonal predictions and a reduction in the maximum single-class error from 9 to 3 in PointNet. In particular, the bi-directional confusion between the confusing pig5 and pig9 is almost eliminated. Experiments show that the introduced local context and bilinear regularization strategy effectively enhances the ability to distinguish fine-grained feature differences and significantly reduces the misclassification rate between similar individuals.

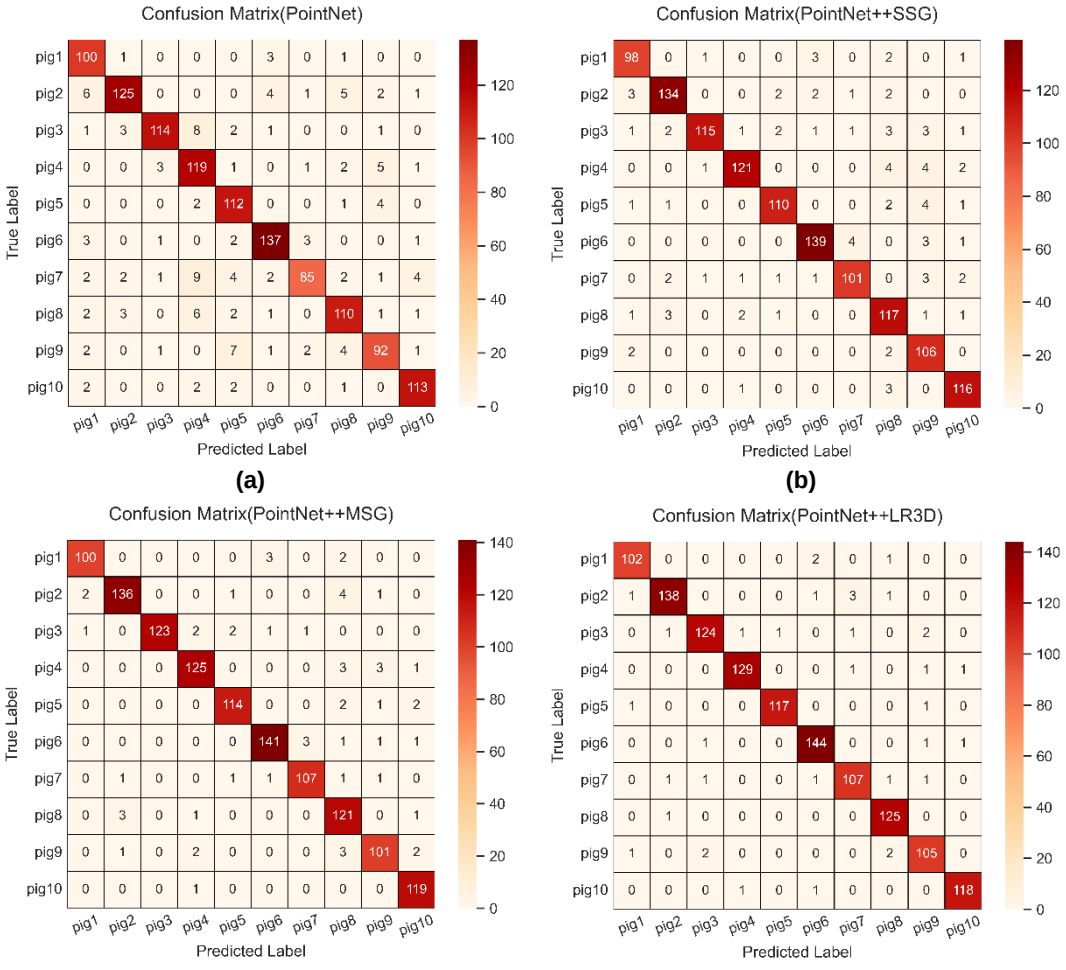


Fig. 10 - Classification confusion matrix for the ten pigs

(a) PointNet classification confusion matrix(b) PointNet++SSG classification confusion matrix(c) PointNet++MSG classification confusion matrix(d) PointNet ++LR3D classification confusion matrix

Discussion

Since the classification performed in this experiment and that performed in the ModelNet40 dataset are fundamentally different, the classification task in ModelNet40 is to classify different objects with significant differences in shape features, while the pig individual identification task performed in this experiment is to classify similar individuals within the same species, which inherently increases the classification difficulty.

To further validate the feasibility of the innovative LR3D bilinear regularization module introduced in this study for individual pig back point cloud identification, extended tests were conducted on the ModelNet40 public point cloud dataset, a classic benchmark for point cloud multi-classification tasks. Testing on this public dataset aims to evaluate the generality and robustness of the proposed method and verify its performance across different point cloud data scenarios.

The test results show that the improved PointNet++ model proposed in this study, by incorporating two LR3D modules, achieves an overall accuracy (OA) and average accuracy (macc) of 93.1% and 91.0% on the ModelNet40 dataset, representing improvements of 2.4% over the original PointNet++ model's 90.7% and 87.6%, respectively. This demonstrates that the proposed improved model enhances classification performance across diverse point cloud data scenarios.

CONCLUSIONS

This study proposes an individual identification method for pig back point cloud data based on the improved PointNet++ model. By introducing local context fusion and a global bilinear regularized LR3D module, the model's ability to extract and classify dynamic pig point cloud features is significantly enhanced. Experimental results demonstrate that the improved model achieves an identification accuracy of 97.11% in the individual identification task, performing the best among all methods. Furthermore, extended testing on the ModelNet40 public dataset further demonstrates the generality and effectiveness of the proposed method.

The LR3D module proposed in this paper significantly enhances the geometric sensitivity and semantic expressiveness of point cloud feature extraction through the local context fusion strategy of dual-stream differential coding and the global bilinear regularization mechanism, providing an efficient solution for analyzing complex 3D scenes. Its modular design gives LR3D excellent flexibility, enabling it to adapt easily to various point cloud processing network architectures, including PointNet++, DGCNN, and other mainstream models. It is seamlessly deployed at different feature extraction stages and CNN layer resolutions. LR3D effectively captures rough geometric structures in the low-resolution layer, such as edges and curvature changes; in the high-resolution layer, its ability to refine local features further enhances semantic differentiation. Experiments show that LR3D significantly improves the model accuracy in the classification task of similar individual identification while maintaining the balance of computational efficiency, demonstrating strong robustness and versatility. Looking ahead, the LR3D module can be widely applied to tasks such as point cloud segmentation and detection, providing a solid feature enhancement framework for 3D vision research.

Compared with traditional physical tags and RFID identification, this study achieves non-contact and high-precision pig individual identification using 3D point cloud technology, which overcomes the problems of fragile tags, the high cost of manual operation, and environmental interference. It provides an efficient solution for individual tracking, health monitoring, and behavioral analysis in fine-tuned farming. However, there are still some limitations in this study: the experimental sample size is small, with only 10 pigs and a single pig breed, and the model can be extended to a larger and multi-breed pig population to verify the generalization ability in the future. In addition, the dynamic point cloud data acquisition is limited by the camera view and the moving speed of the pigs, and the data acquisition process needs to be further optimized. This study provides innovative ideas and technical support for animal individual identification based on point cloud, which has high theoretical value and application prospects.

ACKNOWLEDGEMENT

This work has been supported by the Shanxi Province Basic Research Program (Free Exploration Category) (202303021222066), the ‘Middle-aged and Young Elite Innovative Talent Cultivation Project’ (SXAUKY2024002).

REFERENCES

[1] García, R., Aguilar, J., Toro, M., Pinto, A., & Rodríguez, P. (2020). A systematic literature review on the use of machine learning in precision livestock farming. *Computers and Electronics in Agriculture*, *179*, 105826.

[2] Guo, Y., Wang, H., Hu, Q., Liu, H., Liu, L., & Bennamoun, M. (2020). Deep learning for 3d point clouds: A survey. *IEEE transactions on pattern analysis and machine intelligence*, *43*(12), 4338-4364.

[3] Krampe, C., Ingenbleek, P. T., Niemi, J. K., & Serratosa, J. (2024). Designing precision livestock farming system innovations: A farmer perspective. *Journal of Rural Studies*, *111*, 103397.

[4] Kyaw, P. P., Tin, P., Aikawa, M., Kobayashi, I., & Zin, T. T. (2024). Cow’s Back Surface Segmentation of Point-Cloud Image Using PointNet++ for Individual Identification. In *International Conference on Genetic and Evolutionary Computing.* pp. 199-209. Singapore: Springer Nature Singapore.

[5] Luo, J., Zhang, D., Luo, L., & Yi, T. (2024). PointResNet: A grape bunches point cloud semantic segmentation model based on feature enhancement and improved PointNet++. *Computers and Electronics in Agriculture*, *224*, 109132.

[6] Martínez-Miró, S., Tecles, F., Ramón, M., Escribano, D., Hernández, F., Madrid, J., ... & Cerón, J. J. (2016). Causes, consequences and biomarkers of stress in swine: an update. *BMC veterinary research*, *12*, 1-9.

[7] Neethirajan, S., Scott, S., Mancini, C., Boivin, X., & Strand, E. (2024). Human-computer interactions with farm animals—enhancing welfare through precision livestock farming and artificial intelligence. *Frontiers in Veterinary Science*, *11*, 1490851.

[8] Nong, X., Bai, W., & Liu, G. (2023). Airborne LiDAR point cloud classification using PointNet++ network with full neighborhood features. *Plos one*, *18*(2), e0280346.

[9] Qi, C. R., Su, H., Mo, K., & Guibas, L. J. (2017). Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 652-660).

[10] Qi, C. R., Yi, L., Su, H., & Guibas, L. J. (2017). Pointnet++: Deep hierarchical feature learning on point sets in a metric space. *Advances in neural information processing systems*, *30*.

[11] Ruiz-Garcia, L., & Lunadei, L. (2011). The role of RFID in agriculture: Applications, limitations and challenges. *Computers and electronics in agriculture*, *79*(1), 42-50.

[12] Saleem, M. H., Potgieter, J., & Arif, K. M. (2021). Automation in agriculture by machine and deep learning techniques: A review of recent developments. *Precision Agriculture*, *22*(6), 2053-2091.

[13] Servi, M., Profili, A., Furferi, R., & Volpe, Y. (2024). Comparative Evaluation of Intel RealSense D415, D435i, D455 and Microsoft Azure Kinect DK Sensors for 3D Vision Applications. *IEEE Access*.

[14] Shuai, S., Ling, Y., Shihao, L., Haojie, Z., Xuhong, T., Caixing, L., ... & Hanxing, L. (2020). Research on 3D surface reconstruction and body size measurement of pigs based on multi-view RGB-D cameras. *Computers and electronics in agriculture*, *175*, 105543.

[15] Su, J., Tan, B. E., Jiang, Z., Wu, D., Nyachoti, C. M., Kim, S. W., ... & Wang, J. (2024). Accelerating precision feeding with the internet of things for livestock: From concept to implementation. *Science Bulletin*, *69*(14), 2156-2160.

[16] Vidal, M., Wolf, N., Rosenberg, B., Harris, B. P., & Mathis, A. (2021). Perspectives on individual animal identification from biology and computer vision. *Integrative and comparative biology*, *61*(3), 900-916.

[17] Wang, S., Jiang, H., Qiao, Y., Jiang, S., Lin, H., & Sun, Q. (2022). The research progress of vision-based artificial intelligence in smart pig farming. *Sensors*, *22*(17), 6541.

[18] Yu, S., Liu, X., Tan, Q., Wang, Z., & Zhang, B. (2024). Sensors, systems and algorithms of 3D reconstruction for smart agriculture and precision farming: A review. *Computers and Electronics in Agriculture*, *224*, 109229.

[19] Zhang, K., Ren, W., Luo, W., Lai, W. S., Stenger, B., Yang, M. H., & Li, H. (2022). Deep image deblurring: A survey. *International Journal of Computer Vision*, *130*(9), 2103-2130.

[20] Zhou, H., Li, Q., & Xie, Q. (2023). Individual pig identification using back surface point clouds in 3D vision. *Sensors*, *23*(11), 5156.