

## Original papers

**FSDNET: A features spreading net with density for 3D segmentation in agriculture**Qinghe Liu<sup>a</sup>, Huijun Yang<sup>a,b,c,\*</sup>, Junjie Wei<sup>a</sup>, Yuxuan Zhang<sup>a</sup>, Shuo Yang<sup>a</sup><sup>a</sup> College of Information Engineering, Northwest A&F University, Xinong, Yangling 712100, Shaanxi, China<sup>b</sup> Key Laboratory of Agricultural Internet of Things, Ministry of Agriculture and Rural Affairs, Yangling 712100, Shaanxi, China<sup>c</sup> Shanxi Engineering Research Center of Agricultural Information Intelligent Perception and Analysis, Yangling 712100, Shaanxi, China

## ARTICLE INFO

## ABSTRACT

## Keywords:

Point cloud  
Deep learning  
Fruit segmentation  
Gaussian density  
Strawberry dataset

The accurate segmentation of fruit phenotypes in the field is of great significance for agricultural automation in the 3D scene. Although the existing fruit segmentation based on 3D point cloud has made great progress, in the complex field environment, due to lighting, leaf occlusion, shooting angle and other problems, the point cloud obtained by depth camera often has the problem of multiple voids and discrete points, which seriously affects the accurate segmentation of fruit phenotype. This paper proposes a embedding subnetwork FSDnet based on density-based feature extraction and feature propagation and embeds it in the novel segmentation networks, which effectively improves the segmentation accuracy of the point cloud phenotype in multi-hole and multi-discrete fruits, including (1) The density-based point cloud feature extraction and feature propagation theory is proposed to alleviate the problem of perception degradation in fruit edge point caused by discrete points and holes caused by incomplete point cloud in the agriculture scene. (2) A density-adaptive embedding semantic segmentation framework FSDnet is proposed, and embedding the classical point cloud neural network can significantly improve the segmentation accuracy of the fruit phenotypes with multiple holes and discrete points in the traditional network. (3) This paper made a strawberry dataset and tested the designed new neural network on both strawberry and apple field dataset. After FSDnet is embedded on different novel net, almost all net have been improved. We verified the performance of FSDnet in different density states in agricultural scenarios, mitigated the negative impact of density on segmentation accuracy, proving that it can adapt to different point cloud density in agricultural scenarios in comparison between Gaussian density and other two traditional density schemes, Gaussian density reduces the computational traffic (0.58G) of the network while maintaining similar performance to the other two densities, proving the superiority of assuming a Gaussian density.

**1. Introduction**

With the development of agricultural automation, how to efficiently identify fruit phenotypes and their environmental characteristics and further improve the accuracy of automatic picking has become the core issue of current fruit automatic picking. The main problem is the accurate segmentation of the fruit phenotype data from its environment. At present, between the two main data forms of image and point cloud, point cloud can reflect the phenotype and environmental characteristics of crops more comprehensively, which becomes an emerging research direction.

The traditional point cloud technology is mainly uses LiDAR to get point cloud, but in agricultural scenarios, high planting density and

complex plant characteristics, coupled with the susceptibility to strong light interference, low color recognition of lidar scanning equipment, make it impossible for large-scale applications in agricultural field. With the development of depth cameras and the advancement of point cloud inversion theory, depth cameras based image reconstruction overcomes many shortcomings of traditional 3D scanners and becomes the key means to collect crop point clouds. Although the depth camera is cheap and portable, it can effectively plant feature capture and ensure the reconstruction integrality of plant traits with high accuracy, and it has the prospect of large-scale application in the agricultural field. However, affected by the environment and angle, the fruit point cloud obtained by this method has many discrete points, lacks and holes, which seriously affect the segmentation accuracy of the edge point in the fruit point

\* Corresponding author.

E-mail address: [yhj740225@nwsuaf.edu.cn](mailto:yhj740225@nwsuaf.edu.cn) (H. Yang).

cloud. Therefore, restoring the missing features caused by discrete points and holes is the key for efficient segmentation of fruit phenotypes from environment.

In recent years, point cloud semantic segmentation has made great progress. Mainstream point cloud semantic segmentation schemes have been formed based on projection, discretization and original point cloud. Among them, the original point cloud-based method has become the main solution for the semantic segmentation of agricultural point cloud because of low computational traffic, low equipment requirements and high feasibility. However, the current network is mainly applied to semantic segmentation datasets in industrial and indoor environments (S3DIS (Hightower et al., 2000), ScanNet (Dai et al., 2017), ShapeNet (Chang et al., 2015)), and there is a lack of attention to the processing of multi-view generated point clouds in agricultural scenes, which are often accompanied by holes and discrete points due to the complexity of agricultural scenes. Therefore, it will affect the segmentation effect of the current network structure on the dataset.

Based on this, we propose a embeddable density-based semantic segmentation framework for fruit point clouds: FSDnet (Features Spreading Net with Density). Feature extraction and feature propagation are performed by minimizing the local density of fruit point clouds to improve the semantic segmentation accuracy of field fruits in the case of missing and multiple discrete points. The main contributions are as follows:

A density-based point cloud feature extraction and feature propagation learning method are proposed, in order to alleviate the problem that the point cloud has many holes and multiple discrete points in the agricultural scene, which leads to the decrease of fruit detail perception.

A density feature-based embedding subnetwork FSDnet is designed to implement density-based feature extraction and feature propagation methods, which alleviates the accuracy degradation of fruit phenotype segmentation in traditional frameworks.

A semantic segmentation model is proposed for accurately segmenting fruits under uneven field point cloud density, which provides a new idea for accurately extracting the detailed phenotypic characteristics of field fruits with holes and lacks.

## 2. Related work

### 2.1. Current status of point cloud deep learning

With the progress of point cloud deep learning technology, point cloud semantic segmentation technology has been widely used in the automatic driving, robot operation and virtual reality. hitherto, three mainstream methods have been developed based on projection, discretization and original point cloud.

**Projection-based methods.** In 2017, Jaremo Lawin et al. (2017) first adopted the method of projecting multiple angle views onto a 2D plane. Similarly, Boulch et al. (2017) generated several RGB and depth snapshots of a point cloud using multiple camera positions. However, they cannot solve the loss of accuracy from occlusion and viewpoint selection. The projection-based method solves the problem of difficult segmentation caused by the disorder of the point cloud. However, with the increase of the number of point clouds, the projection method will seriously consume cost.

**Discretization-based methods.** In 2016, Huang and You (2016) first divided the point cloud into a set of voxel grids, and then fed these intermediate data into a full 3DCNN for voxel segmentation. Meng et al. (2019) introduced a kernel-based interpolating variational autoencoder architecture. Tchapmi et al. (2017) proposed SEGCloud to achieve fine-grained and globally consistent semantic segmentation. The granularity of such methods and boundary artifacts causes point cloud division in loss of features. And due to large computation voxelization, voxel based network is more computational complexity than point based methods.

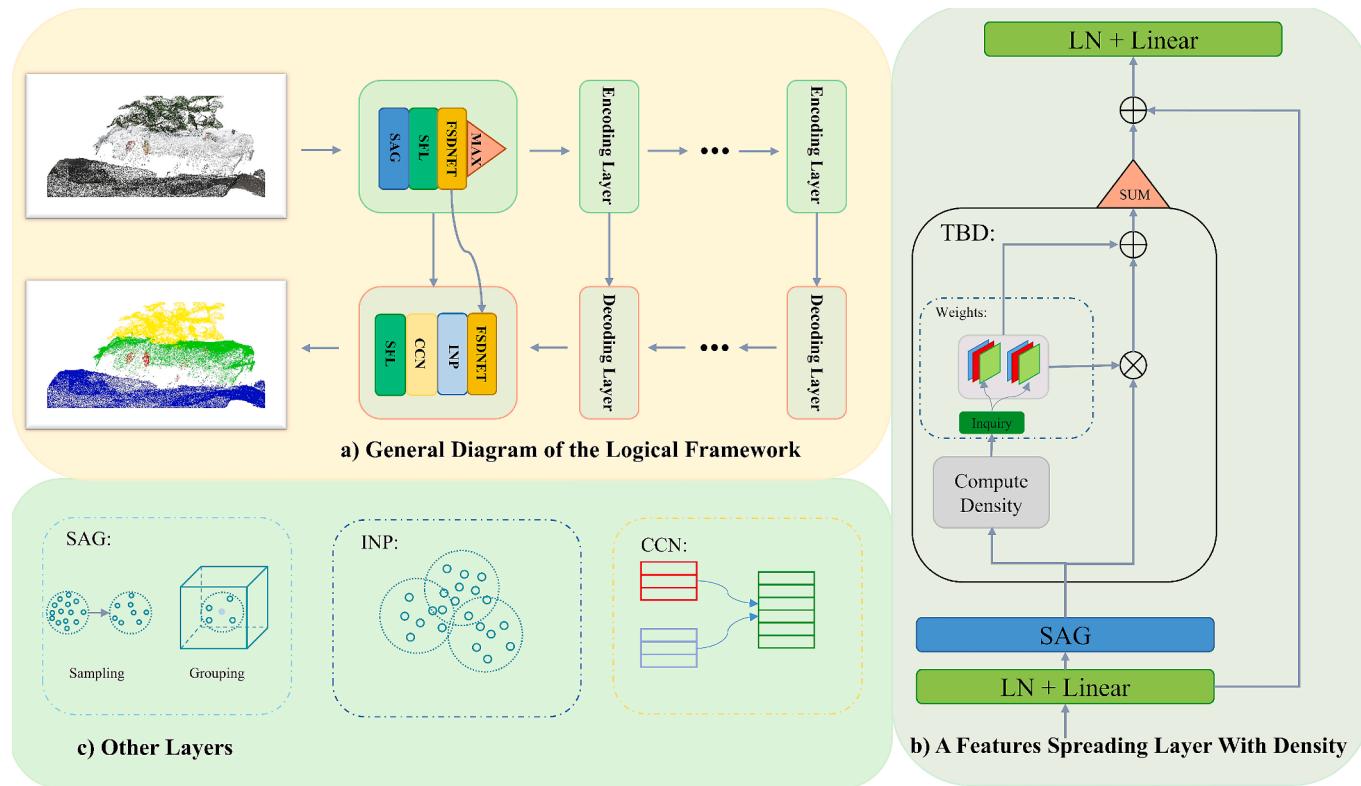
**Methods based on raw point clouds.** In 2017, pointnet (Qi et al., 2017) made pioneering work on this method. Encoding, pooling,

extracting features, and decoding each point through shared mlp can theoretically effectively deal with the disordered structure of point clouds, but because of the loss of features due to pooling, and the lack of capture of local spatial geometric features, it failed to achieve the desired effect. After that, the point cloud semantic segmentation network was mostly improved on the basis of pointnet, and developed the following four mainstream improvement ideas.

- (1) Improvements based on the MLP network. To enhance the capture of local geometric features, Ni et al. (2020) proposed the pointnet++ by grouping and sampling point clouds for feature extraction and propagating with linear interpolation. Zhao et al. (2019) proposed PointWeb, which employs the AFA module to capture the relationships among neighboring points during feature extraction, yielding promising results. Nonetheless, the approach is constrained by the exclusive use of MLP for learning extracted features, making it challenging to enhance the accuracy further.
- (2) Improvements based on RNN. Huang et al. (2018a) proposed a lightweight local modeling and used slice pooling layers to convert unordered points into ordered vector sequences. To alleviate the problems caused by static pooling operations, Huang et al. (2018b) proposed a dynamic aggregation network (DAR-Net) to handle both global scene complexity and local geometric features. The above methods improve the learning ability of local features, but they increase the computational cost.
- (3) Improvements based on graph neural networks. Landrieu and Boussaha (2019) represented a point cloud as a set of interrelated simple shapes and superpoints and used a superpoint graph to capture structural and contextual information. Zhiheng and Ning (2019) proposed a Pyramid Net based on Graph Embedding Module (GEM) and Pyramid Attention Network (PAN), but the construction of super-point graphs for point clouds remains to be studied.
- (4) Improvements Based on Point Convolutions. Thomas et al. (2019) proposed KPConv to achieve the approximation of spatial convolution, but its preset weights make it impossible to achieve the optimal weight combination and lose the flexibility of convolution to dynamically modify the weights. Xu et al. (2021) used scorenet to learn point positions to flexibly assemble kernel weights. During backpropagation, the original gradient of a convolution is divided by multiple scores learned by scorenet, so the sensitivity of weight modification is weakened.

### 2.2. Related work on density estimation

Wu et al. (2019) proposed PointConv, a method that utilizes a spatial convolution design based on density and position weights. However, their approach employs a ternary Gaussian density estimation, which significantly deviates from the actual density. Li et al. (2020) introduced Density-aware Convolution, which uses a multivariate Gaussian kernel and achieves more accurate density estimation through smoothing factors and normalization. Nonetheless, the smoothing factors and normalization constants are learned using a multilayer perceptron (MLP), adding additional overhead. Mao et al. (2019) proposed an interpolation-based spatial convolution approach that employs a simplified multivariate Gaussian density kernel for density-based interpolation. However, they neglected the covariance matrix when designing the Gaussian kernel, leading to reduced accuracy in density estimation. But due to the large amount of covariance matrix in the actual calculation, the use of complete three-dimensional normal density will lead to the overload of computing resources, and the calculation period will be too long or even incomputable. Therefore, we propose a method that uses three one-dimensional normal densities to be mapped to integer intervals by hash functions and then used as lookup indexes for specific parameter tables, so as to reduce the impact of ignoring



**Fig. 1.** (a) The main logic of our work.sag:sampling and grouping.max:max pooling.inp:interpolation.sfl: Specific Layer, the core layer of the network we have embedded. FSDnet: the core layer of our work. (b) The main logic of core layer. (c) Non-keyframe layer.

independence and enable specific parameter tables to learn specific parameter preferences.

### 2.3. Point cloud deep learning in crop segmentation

In recent years, thanks to the rapid development of 3D point cloud in the field of computer vision and the spatial learning of point cloud semantic segmentation in the industrial field, scholars have tried to apply point cloud semantic segmentation in the agricultural field. Li et al. (2022) developed the DeepSeg3DMaize system based on Pointnet, which provided a reference for the automated analysis of 3D phenotypic characteristics at the plant individual level. Chen et al. (2021) integrated and improved the local feature aggregation module in RandLA-Net to achieve 3D point cloud semantic segmentation of large-scale structured agricultural scenes. Jayakumari et al. (2021) improved the random sampling scheme of pointnet by segmentation of cabbage, tomato and eggplant. Hu et al. (2021) adopted the HSV spatial color enhancement algorithm on pointnet++ to achieve semantic segmentation of point clouds of rapeseed in rapeseed fields. Yu et al. (2022) added a 3DSTN spatial variation network to align point clouds on pointnet and designed a pyramid-shaped pooling module for feature extraction, realizing the segmentation of apples, pears and lemons. Since the research of point cloud semantic segmentation in the agricultural field has just started, it generally lags behind the development of point cloud deep learning. At the same time, in the field environment with high planting density and complex and diverse plant traits, depth cameras can capture complex plant traits with low price, which has the prospect of large-scale application.

As previously stated, in real-world agricultural production, depth camera image synthesis datasets are suited for complex field planting scenarios. However, factors such as shooting angles may lead to the presence of holes and discrete points. We intend to devote increased attention to addressing these issues in future research. Therefore, the segmentation of the edge points of the holes on image synthesis point

cloud has become an urgent problem to be solved.

### 3. The design of features spreading net with density

For data sets with multiple discrete points and multiple holes in agricultural scenarios, the density of fruit point clouds at the edges of holes is low, while the density on the rays along the edge to the center of the dataset increases sharply. In this case, the feature points of fruit points are nearly all interior dense point. In the traditional method, since the interior points account for the majority of the feature points, the segmentation probability of the fruit point tends to be consistent with the interior dense points instead of the low dense exterior points. However, fruit edge points, especially the points at the edge of the holes caused by uncompleted point cloud, their categories are often inconsistent with interior points, thus eventually leading to segmentation errors.

We propose a density-based feature extraction and propagation method to solve the above problems in fruit segmentation. For the method based on the original point cloud, there are often two actions of feature extraction and feature propagation. However, due to the disorder of the point cloud structure, simple feature extraction and feature propagation will cause loss of features in the fruit point cloud with holes and discrete points. Therefore, we add the density calculation module and the density weight to correct the feature, improve the efficiency of feature extraction and propagation, and realize the segmentation of fruit edge points. We call the above network architecture FSDnet (Features Spreading Net with Density) Fig. 1.

#### 3.1. Design of density calculation scheme

The traditional calculation of point cloud density is mainly in the point cloud preprocessing stage, and the algorithm is complex and computationally expensive. In this paper, a Gaussian density is proposed to replace the traditional density estimation method.

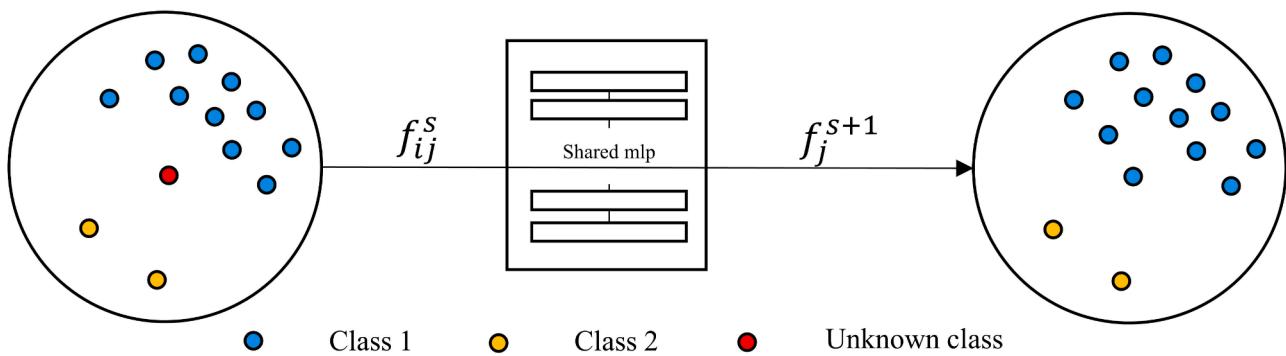


Fig. 2. Feature extraction of mlp network.

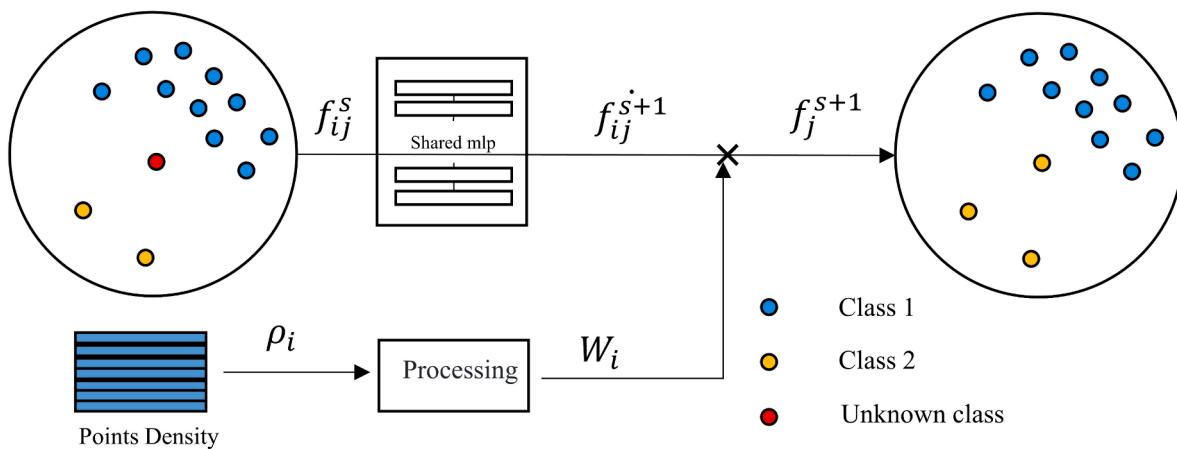


Fig. 3. Density-based feature extraction.

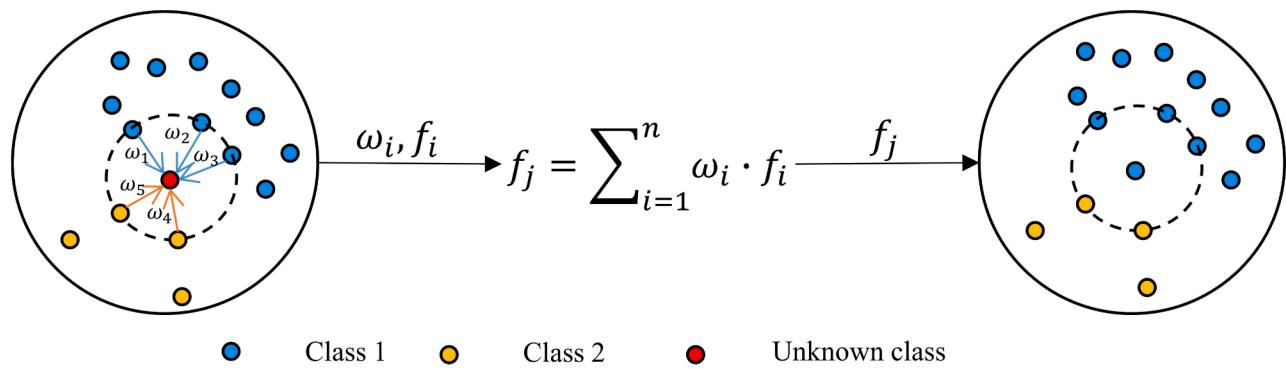


Fig. 4. Linear interpolation.

### 3.1.1. Traditional density prediction scheme

- (1) For each point, by KNN sampling process, suppose k points are sampled, and the reciprocal of the distance of the farthest sampling point is selected as the density estimation of the point, as shown in formula (1).

$$\rho_i = \frac{1}{\max\{\text{dist}_{ij}, j = 0, \dots, k\}} \quad (1)$$

- (2) On the basis of the number of sampling points above, take the reciprocal of the average value of the distance sum from the K sampling points as the density estimation, as shown in formula (2).

$$\rho_i = \frac{1}{\text{mean}\{\text{dist}_{ij}, j = 0, \dots, k\}} \quad (2)$$

### 3.1.2. Gaussian density estimation scheme

Although the above two schemes can better present the density distribution of fruit point clouds, they have a large amount of calculation and are difficult to popularize on a large scale. In order to reduce the computational traffic, we adopt Gaussian density as the core density estimation scheme.

Below we give proof that the distribution probability of the point cloud conforms to a Gaussian distribution in the fruit point cloud scene with large number of points:

We assume that a certain fruit point cloud set is  $P = \{p_i = (x_{i1}, x_{i2},$

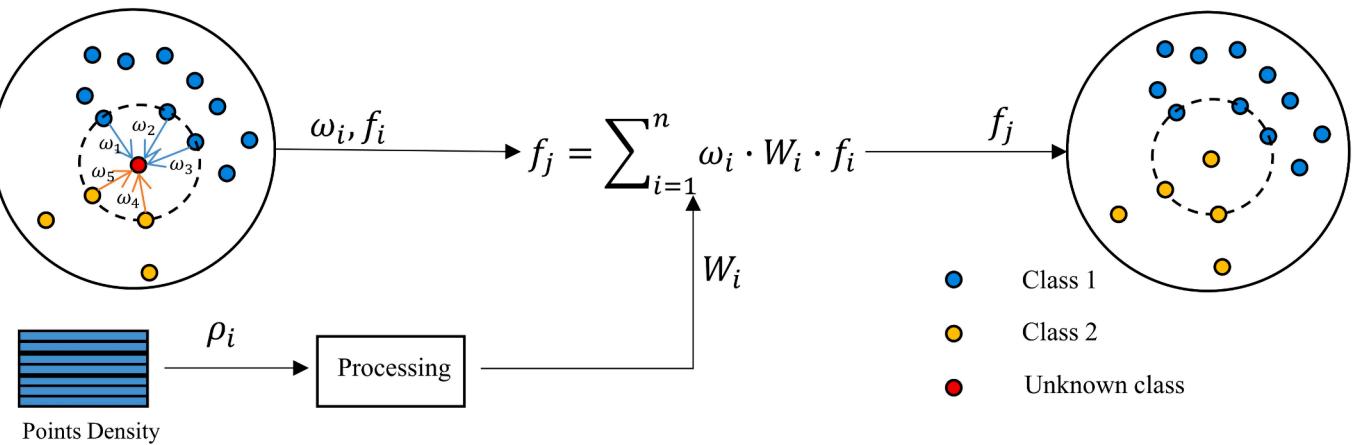
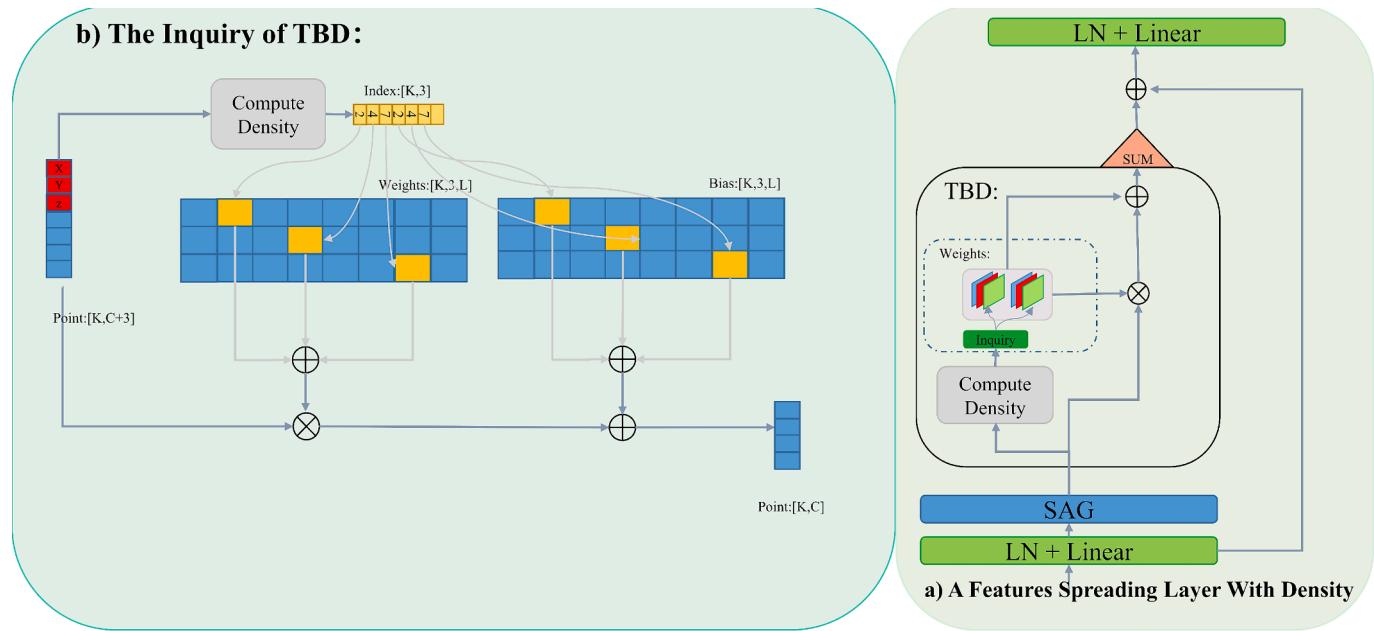


Fig. 5. Density-based interpolation.



**Fig. 6.** Elaborating on Density Weighted Learning Networks. (a) A Features Spreading Layer With Density. LN + Linear: Serially passing through a Linear and a LayerNorm. SAG:Sampling and Grouping as the premise of computational density.TBD: Table Bank of Density, provide the weights based on the index of density. (b) The process of inquiry in TBD.

$x_{i3})| i = 0, 1, 2, 3, \dots, n\}$ , then  $\forall i \in [1, n], \exists F$ , there is  $d_i = F(x_{i1}, x_{i2}, x_{i3})(d_i)$  for the probability density, and  $(x_{i1}, x_{i2}, x_{i3})$  is the coordinates of a point set relative to a central point.), obviously the distribution is independent of each other when the fruit point cloud  $p_i$  is limited, and the mathematical expectation of each dimension is limited and cooperative in the case of no repeated points. Assuming the mathematical expectation of each dimension is  $E(x_r) = u_r, r = 1, 2, 3$ ; The covariance matrix is  $Cov(x_{i1}, x_{i2}, x_{i3}) = (\sigma_{kj})(k, j = 1, 2, 3), i = 1, 2, \dots, n$ , we can deduce formula (3) from Levy-Lindberg central limit law.

That is, for a certain fruit point cloud scene, when the number of point clouds is large enough,  $p_i$  obeys the normal distribution of the three-dimensional vector.

However, due to the need to calculate the covariance matrix of the 3D normal density, the computational resources are too large and the calculation period is too long, which makes it difficult to apply the 3D normal density in real-time training or inference. Therefore, we use the Gaussian probability density of points to replace the traditional density

$$\lim_{n \rightarrow \infty} F_n(x_1, x_2, x_3) = \lim_{n \rightarrow \infty} \left\{ \frac{\sum_{i=1}^n x_{ri} - n\mu_r}{\sqrt{n}(\sigma_{kj})} \leq y_r, r = 1, 2, 3 \right\} = \frac{1}{(\sqrt{2\pi})^3} \int_{-\infty}^{y_3} \int_{-\infty}^{y_2} \int_{-\infty}^{y_1} e^{-\frac{1}{2} \sum_{i=1}^3 \sum_{j=1}^3 \sigma_{ij} u_1 u_2 u_3} du_1 du_2 du_3 \# \quad (3)$$

for calculation (detailed implementation in Section 3.4). Replacing the



**Fig. 7.** Partial point cloud scene display of strawberry dataset.

**Table 1**  
The number of point cloud (m: millions).

DatasetCategory	Floor 0	Base 1	Plant 2	Fruit 3	Total
Training set	263.9 m	128.6 m	240.8 m	16.5 m	649.9 m
Test set	64.5 m	42.7 m	42.1 m	1.4 m	150.8 m
Total	328.4 m	171.4 m	282.9 m	17.9 m	800.8 m

traditional density estimation scheme with Gaussian density can reduce the calculation and improve the feature recognition accuracy. We will design Gaussian density into the below two feature processing schemes and analyze its better performance.

### 3.2. Density-based feature extraction

According to the analysis of the characteristics of the data set, in the segmentation of edge points, the contribution of point sets with low density should be appropriately higher, and that with high density should be lower because of similar characteristics. This view point is worthy of discussion. Let the fruit point cloud set be  $P = \{p_i = (x_i, y_i, z_i) | i = 1, 2, 3, \dots, n\}$ , for any fruit point cloud  $p_i$ , its corresponding feature point set is  $p_{iA} = \{p_{ij} | j = 1, 2, \dots, m\}$ , we also regard the input feature points as unprocessed features, then there are  $F = \{f_{ij}^s | s = 0, 1, \dots, c; i = 1, \dots, n; j = 1, \dots, m\}$ , where  $s$  means after the  $s$ -th layer network operations. Each feature corresponds to a point density, then the feature point density set can be expressed as  $d_{iA} = \{d_{ij} | j = 1, 2, \dots, k\}$ , and the feature extraction operation can be

expressed as formula (4):

$$f_i^{s+1} = \Lambda \left\{ K(p_i, f_{ij}^s) | i = 1, \dots, n; j = 1, \dots, m; s = 0, 1, \dots, c \right\} \quad (4)$$

where  $K$  is the convolution operation and  $\Lambda$  the pooling operation,  $d_{ij}$  is calculated in Eq. (10).

In the features  $f_{ij}^s$  of the characteristic points input by the  $s$ -th layer of convolution of edge points, we set the probability of feature points with high density to appear as  $F_1$ , and that with low density  $F_2$ , obviously,  $F_1 > F_2$ .

For a point  $p_i$  to be classified, it can be expressed as formula (5):

$$f_i^{s+1} = \Lambda \left\{ K_r(p_i, f_{ij}^s) \bullet W_i | i = 1, \dots, n; j = 1, \dots, m; s = 0, 1, \dots, c \right\} \quad (5)$$

We will use weight calculated from density (Section 4.4) to attenuate the contribution of large density points to extract edge point features. On the other hand, in fruit point clouds with similar densities, the weights of all points are similar, so it will not cause a drop in the segmentation accuracy of internal points.

We assume that the fruit point cloud is divided into two classes (class 1, class 2), then for an unknown point (its true class is 2), its features are often from surrounding points (feature points) and the constituted local parts determined by geometric features. In Fig. 2, the points of class 1 around the unknown point are high-density points, and the points of class 2 are low-density points. The easiest way to get unknown point features is to sample a certain number of nearby points around and learn through a mlp to get their point features. Due to the disorder of the input point cloud, its mlp must use the same parameters for each feature point to be extracted (using  $1 \times 1$  convolution), so it is impossible to use different weights for different points. This method will cause low-



**Fig. 8.** Apple dataset sample point cloud.

**Table 2**  
The IOU of strawberry dataset network training (%).

ModelCategory	Floor	Base	Plant	Fruit	mIOU
Pointnet	61	77	66	0.0	51
Pointnet2	71	79	89	38	69
PointWeb	73	82	91	53	74.2
PACconv	76	80	90	47	76.4
KPConv	72	83	93	51	77.0
PointConv	74	80	88	45	73.7
Pointnet + FSDnet(Gaussian)	65	85	78	5	58
Pointnet2 + FSDnet(Gaussian)	70	85	89	43	72
PointWeb + FSDnet(Gaussian)	73	81	93	59	77.2
PACconv + FSDnet(Gaussian)	77	82	91	44	80.9
KPConv + FSDnet(Gaussian)	75	82	90	47	76.8

density points and high-density points to adopt a consistent linear transformation method when extracting features, resulting in the point features to be segmented more similar in high-density points. The schematic process is shown in Fig. 2.

In order to flexibly handle feature points with different density and weaken the contribution of high-density points, we learn the density of feature points through a series of processes to get their density weights, and then redistribute the contributions of different density points to unknown points through the density weights to correct high-density points as shown in Fig. 3.

### 3.3. Density-based feature propagation

In the feature propagation stage, we replace traditional linear interpolation model with density-based interpolation. The feature set of sampling points obtained from the feature extraction stage is  $F = \{f_i | i = 1, \dots, k\}$ . The unknown feature set of the global point is  $\bar{F} = \{f_j | j = 1, \dots, n\}, n > k$ ; For  $\forall j \in [1, n]$ , there is  $f_j = \sum_{i=1}^k f_i \bullet \omega_i$ , where  $\omega_i$  is the contribution weight. Euclidean distance is used to quantify the weight formula (6):

$$\omega_i = \frac{\text{dist}_i}{\sum_{r=1}^n \text{dist}_r} \quad (6)$$

Then the unknown feature of each global point can be obtained according to formula (7):

$$f_j = \sum_{i=1}^n \omega_i \bullet f_i \quad (7)$$

The traditional linear interpolation measures the weight by distance, which is effective when the fruit point cloud density distribution is constant. However, when the distance weights are the same, the high-density sampling points in class 1 and the low-density sampling points in class 2 are interpolated to the low-density unknown points (the true labels are still class 2) at the same time, the distance weights  $\omega_i (i = 1, 2, 3, 4, 5)$  are the same, then the high-density points contribute greatly due to their large number. the features in low density point after interpolation are more similar to the high density point, as shown in Fig. 4.

We add the weight calculated from density in the interpolation stage, then the density-based interpolation can be expressed as formula (8).

$$f_j = \sum_{i=1}^n \omega_i \hat{A} \cdot W_i \bullet f_i \quad (8)$$

Through the density optimization, when different density sampling points interpolate to low-density unknown points, these weights are able to flexibly reflect the contribution of different density points, and finally the low-density unknown points get consistent labels, as shown in Fig. 5.

### 3.4. Features spreading net with density

In the design of Features Spreading Net with Density, since the cost of calculating the covariance matrix and variance matrix of the three-dimensional normal distribution is too high, we use three one-dimensional independent normal distributions as a simplification. In order to reduce the impact of its hypothetical three-dimensional independence on the specific implementation, we do not directly use it in the network training, but map it into an integer interval through a hash function, which is further used as the query index of Table Bank of Density, so that FSDnet can learn a special preference for certain weights. The specific modules are implemented as follows:

**LN + Linear:** We first define the input approach of features. First, the features are transformed through a linear layer, which make it into the new feature space. Then, for preventing the gradient explosion or vanishing, we have employed the LayerNorm, compared with BatchNorm, which could decline the introduction of noise, facilitating faster convergence of the network.

**SAG (Sampling And Grouping):** Distinguished from the modules in the backbone network, for defining the perceptual field size of sampling in order to determine the index of density easily, the measure of grouping is “ballquery” instead of “knnquery”. Therefore, we could define the vision size of sampling through the radius.

**Compute Density:** To reduce the computational complexity further, we could consider 3-dimensional coordinates to be independently

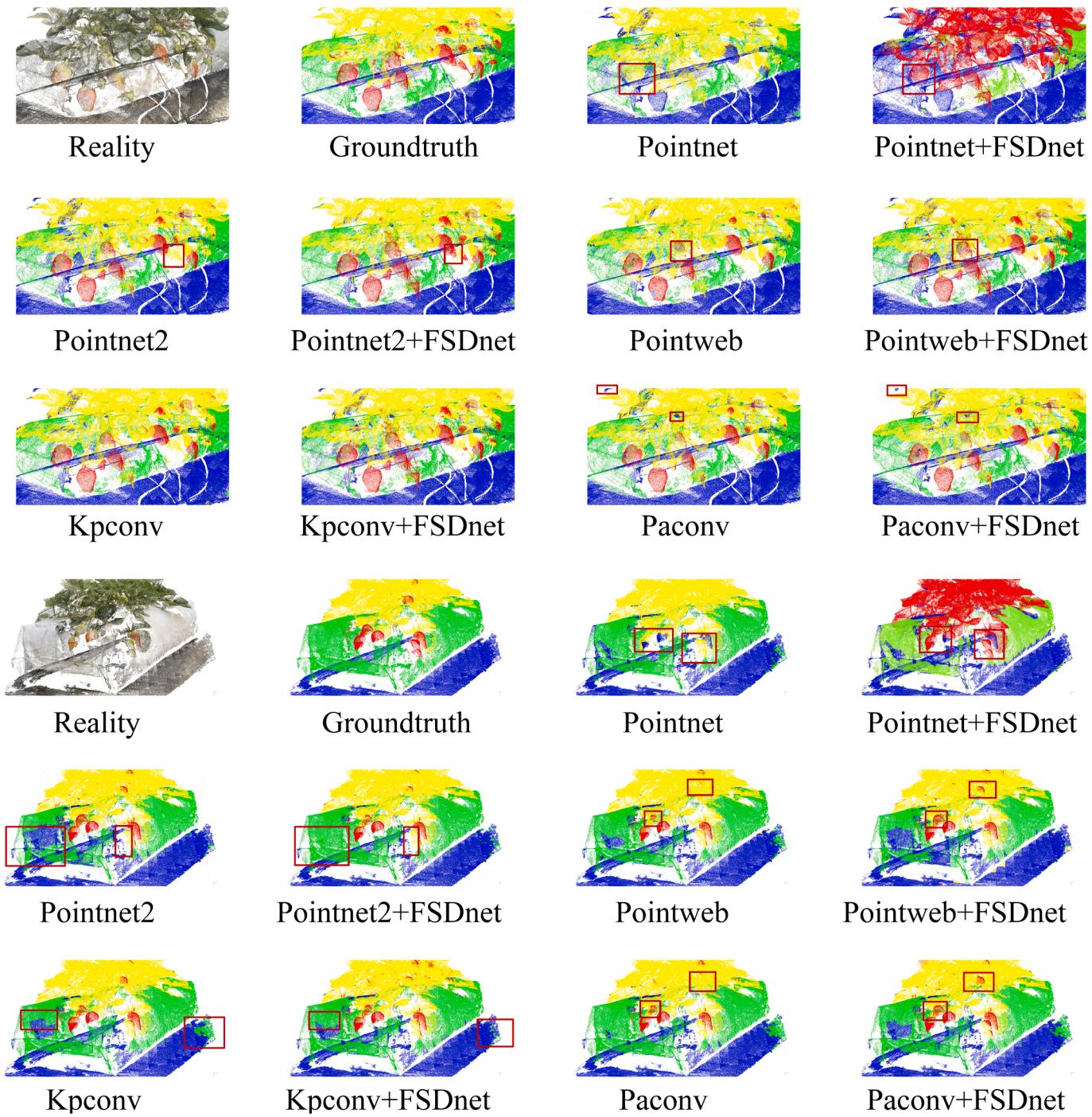


Fig. 9. Visual comparison on strawberry dataset.

distributed under the condition of acceptable error. In order to reduce the negative impact of the assumption of the independence of the three one-dimensional normal distributions, we do not directly involve the three one-dimensional normal densities in the operation, but instead use a hash function to map them to an integer interval and then use them as a lookup index for a specific parameter table (see Fig. 6), which can be expressed as  $\text{idx}_x = \text{hash}(g(x_i))$ ,  $\text{idx}_y = \text{hash}(g(y_i))$ ,  $\text{idx}_z = \text{hash}(g(z_i))$ . (9) where  $g$  denotes the one-dimensional Gaussian density operator. The hash function implementation is described in inquiry.

**Weights:** For each feature dimension, we designed a  $3*L$  size matrix to store the weights corresponding to the density on each coordinate dimension. Thus, its capacity is  $C*3*L$ , in comparison with regular

measure whose capacity is round  $C*S*S*4$  ( $L \cong 3S, S > 3$ ).

**Inquiry:** That's include Query and Calculation. In Query phase, we compute index with the help of density in 3-dimensional coordinates, that could be expressed as " $\text{idx}_{j,x} = \lceil g(x_i)_j * L \rceil$ " where  $\text{idx}_{j,x}$  is the density index in the x direction for the j-th features dimension. In Calculation phase, We have adopted a more modern approach that combines density weighting with features as:

$$f_j = \sum_{i=1}^n W_i f_i + B_i \quad (10)$$

where  $\mathbf{W}_i = L_{1,x}[\text{idx}_{i,x}] + L_{1,y}[\text{idx}_{i,y}] + L_{1,z}[\text{idx}_{i,z}]$ ,  $\mathbf{B}_i = L_{2,x}[\text{idx}_{i,x}] + L_{2,y}[\text{idx}_{i,y}] + L_{2,z}[\text{idx}_{i,z}]$ .

**Table 3**

The IOU of apple dataset network training (%).

Model\Category	Branch	Leaf	Fruit	MIOU
Pointnet	21	10	76	36
Pointnet2	89	73	95	91
PointWeb	87	92	95	92
PACConv	91	95	95	94
KPConv	88	93	96	93
PointConv	90	86	94	92
Pointnet + FSDnet(Gaussian)	22	20	75	41
Pointnet2 + FSDnet(Gaussian)	93	79	95	93
PointWeb + FSDnet(Gaussian)	90	96	94	93
PACConv + FSDnet(Gaussian)	93	97	96	95
KPConv + FSDnet(Gaussian)	87	92	95	91

In the specific implementation, we adopted two weights, W and B, as density weights, which allows the feature transformation to map to a larger feature space. Then, by embedding this calculation process before feature extraction and feature propagation, we can accomplish the tasks of Formula 5 and Formula 8.

It should be noted that for the two traditional density calculation schemes in [Section 3.1.1](#), we only need to take their 3D coordinate vectors as input and perform density calculations separately. In the subsequent experimental section, we embedded our proposed module into well-established network architectures, including PointNet, PointNet++, PointWeb, PACConv, and KPConv. This integration aims to demonstrate the efficacy and versatility of the module in enhancing the performance of these networks.

## 4. Experiment

### 4.1. Data set preparation

#### 4.1.1. Strawberry scene point cloud dataset

In order to verify the validity of our theory and expand the application of point cloud deep learning in agriculture, we choose strawberry as our experimental plant. In the field we use deep camera to acquire images of strawberry plants from different angles, about 60 images are taken per strawberry plants on average, then we use MVS algorithms to reconstruct point clouds from these images. Due to the presence of discrete points, We preprocess the reconstructed point cloud to remove most of the discrete points. The reconstruction point clouds are dense, containing around 4 to 6million points per strawberry plants. The reconstruction scene is a dense point cloud with the field's floor, strawberry's base, strawberry plants and fruits. In order to improve the efficiency of model training, we downsample the point cloud. In the classification, the label of floor is set to 0, the label of base is set to 1, the label of plants is set to 2, the label of fruit is set to 3. At last these labeled point clouds are respectively encapsulated in the h5 files. Compared with other crops, strawberry has a short growth cycle, short plants, and dense branches and leaves. The fruit grows among the branches and leaves, which is severely blocked and difficult to photograph. Therefore, in the process of capturing an image containing strawberry fruit, it is inevitable to cause the angle deviation of the captured image and the deficiency of the plant, and then there will be many discrete points, holes, and uneven density when synthesizing the point cloud. Therefore, it is appropriate to choose strawberries as our experimental data to verify the ability of the network.

In the strawberry experimental field in Yangling City, Shaanxi Province, the strawberry point cloud was photographed and labeled by the depth camera from multiple perspectives, and 30 sets of point cloud scenes in the training and validation set and 7 sets of point cloud scenes in the test set were obtained, shown in [Fig. 7](#). The specific point cloud data are shown in [Table 1](#).

### 4.1.2. Apple dataset

In order to test the generalization performance of proposed FSDnet, we choose the Apple dataset ([Yu et al., 2022](#)) for test. Apple dataset is also a dataset composed of images from multiple angles. This data set is only composed of three types of objects: branches, leaves, and fruits. When capturing images, due to the small number of pictures taken, the features of the synthesized data set are more sparse. Therefore, the segmented fruit phenotype is very different from that of strawberries, An example of this is shown in [Fig. 8](#).

It can be clearly seen from the [Fig. 8](#) that the average density of the Apple data set is lower, but there are almost no discrete points and holes, and all points are almost within a regular range, which has a lot to do with the artificial splicing model.

### 4.2. Experimental comparison of strawberry dataset

The experiment of embedding FSDNet on different networks is first carried out on the strawberry dataset. The experimental results show that the contribution of similar features originally concentrated in the scene is weakened by the density network, and the sparse and discrete features are enhanced, thereby it improves the performance of the original network, especially the segmentation accuracy of fruits. The specific data is shown in [Table 2](#).

It can be seen from the table that the mIOU of Pointnet is increased 7 % by improving the existing classic network structure with the FSDnet module, and for pointnet that lacks overall perception of the scene, after adding FSDnet, the iou (5 %) of the segmentation fruit is a Great improvement. For Pointnet++, mIOU has increased by 2 %, and all IOUs except the floor have been improved to a certain extent. Then, the model has the potential to enhance the accuracy of nearly all advanced models except KPConv. We figure out that the ineffectiveness of embedded FSDnet can be attributed to the excessive number of parameters in KPConv that do not align with the complexity of the dataset. The visualization and training curves are shown in [Fig. 9](#). Blue represents floor, green represent base, yellow represents plants, red represents fruit.

### 4.3. Generalization performance test

In order to verify that FSDnet has universal performance in agricultural fruit segmentation, we conducted a generalization experiment on the Apple Dataset, and extracted 16 groups of the apple segmentation data set used in LFPnet for experiments. The results are shown in [Table 3](#):

As shown in [Table 3](#), we adopted the FSDnet module in Pointnet and Pointnet2. mIOU increased by 5 % and 2 %. And the IOU of most classifications increased. In the segmentation of the edge points in [Fig. 10](#), after adding the FSDnet module, the segmentation of edge parts such as leaf and branch is more accurate, which shows the effectiveness of our improvement. Similar to the results on the strawberry dataset, the embedded FSDnet results of all advanced models, except KPConv, are superior. The visualization and training curves are shown in [Fig. 10](#). Blue represents branch,green represents leaf,red represents fruit.

### 4.4. Comparative experiment of different density data sets

In order to verify that FSDnet has strong performance under different densities, we have performed different subsampling on the two data sets, and the sampling distance unit is defined as equation (11).

$$d_{unit} = \frac{|d_i - d_j|_{max}}{N} \# \quad (11)$$

where  $d_{unit}$  is the unit length,  $|d_i - d_j|_{max}$  is the maximum distance between any two points in the point cloud, and  $N$  is the total number of point clouds.

On this basis, we use uniform sampling and use a fixed unit value as

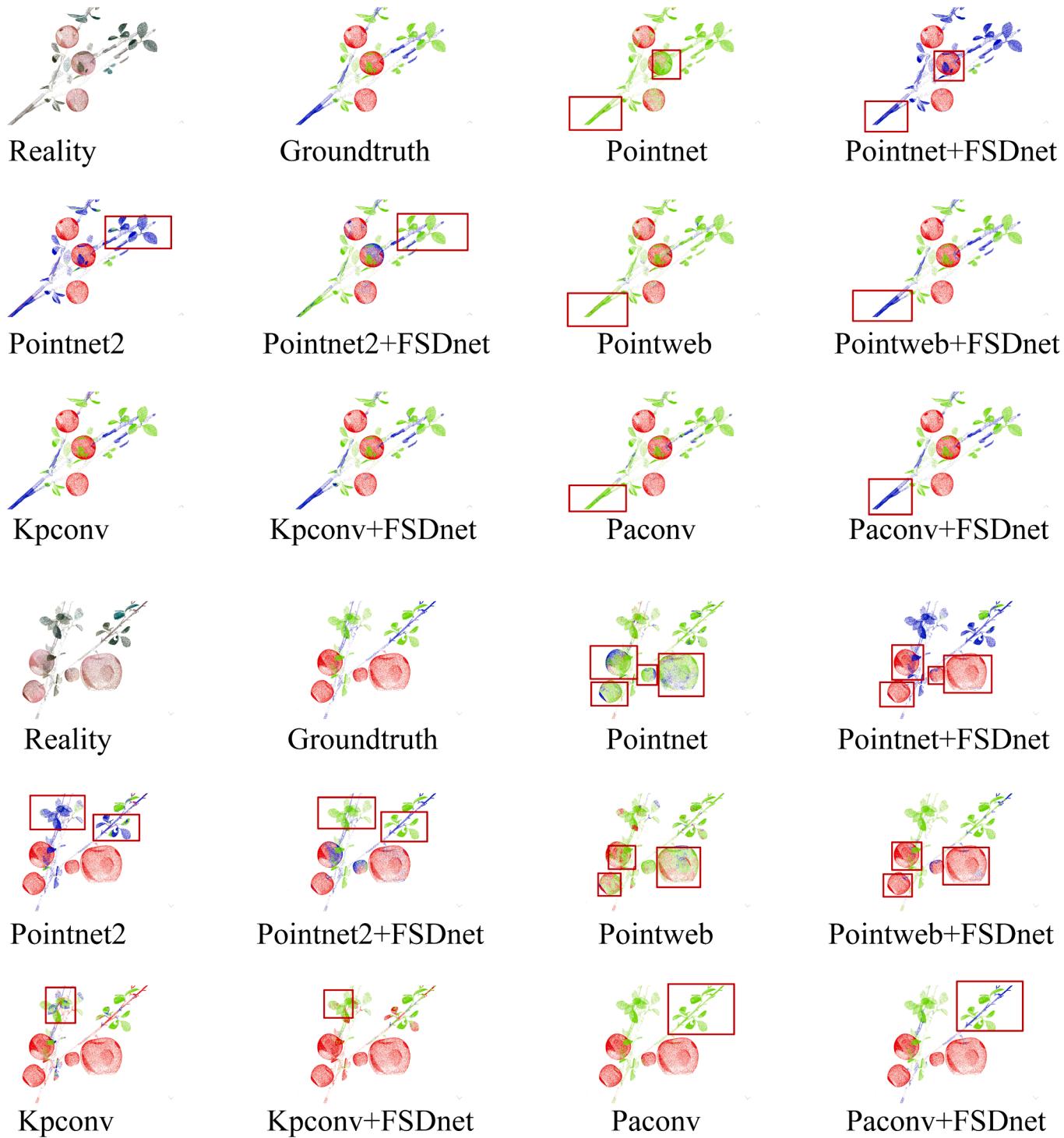


Fig. 10. Visual comparison on apple dataset.

**Table 4**

Accuracy of network segmentation under different densities of apple dataset (%).

Model\Sampling radius	0.001	0.01	0.03
Pointnet	63	41	19
Pointnet2	86	81	78
Pointnet + FSDnet(Gaussian)	71	59	27
Pointnet2 + FSDnet(Gaussian)	88	84	79

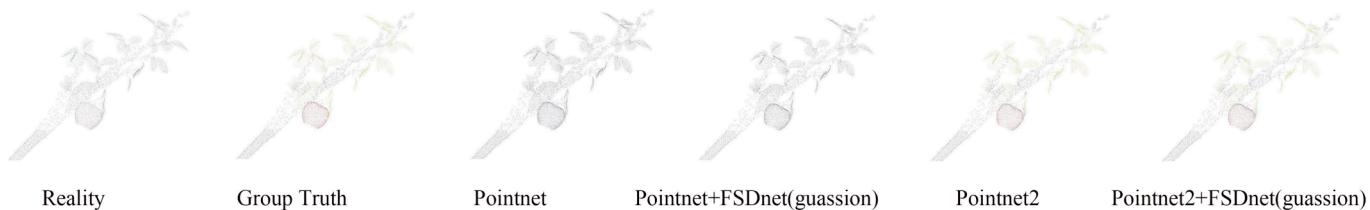
the sampling radius to subsample the strawberry and apple datasets.

#### 4.4.1. Apple dataset density comparison experiment

We use the apple data set with a sampling radius of 0.001, 0.01 and 0.03 unit lengths respectively, and conduct a split test. The accuracy is as Table 4:

It can be seen from the above table that the segmentation accuracy of all models almost decreases with the decrease of the density, but before comparing the embedded FSDnet, the reduction of the segmentation accuracy is significantly reduced after the embedded FSDnet. After

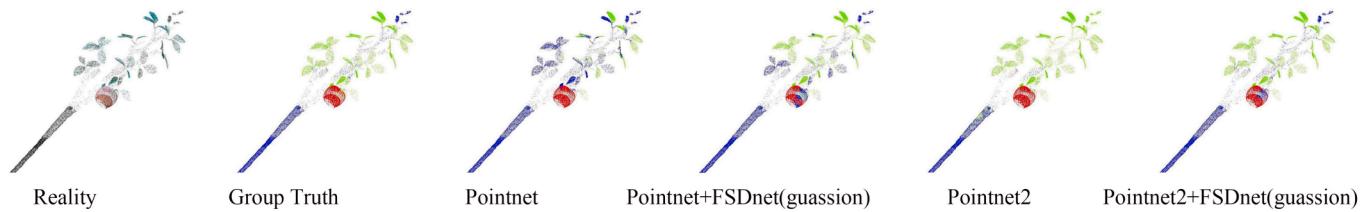
sampling distance:0.03



sampling distance:0.01



sampling distance:0.001



**Fig. 11.** Apple dataset different density results visualization atlas.

**Table 5**  
Accuracy of network segmentation under different densities of strawberry dataset(%).

Model\Sampling radius	0.001	0.01	0.05
Pointnet	65	54	23
Pointnet2	85	73	60
Pointnet + FSDnet(Gaussian)	73	58	34
Pointnet2 + FSDnet(Gaussian)	89	80	62

Pointnet is embedded with FSDnet, the accuracy rate increases by 8 %, 18 % and 8 % respectively under different density conditions; After Pointnet2 is embedded with FSDnet, the accuracy rate increases by 2 %, 3 % and 1 %, respectively, which shows that the density is in a certain range. Inner helps to learn sparse features. Its segmentation visualization is as Fig. 11:

#### 4.4.2. Strawberry dataset density comparison experiment

Like the apple dataset, we subsample the strawberry dataset with a unit radius of 0.001, 0.01 and 0.05, and perform a split test. Since the strawberry point cloud is relatively dense, the maximum sampling radius is larger than that of the apple dataset, and it still performs well. The split test is performed, and the accuracy rate is as Table 5:

It can also be clearly seen that after the FSDnet is embedded, the segmentation accuracy rate decreases significantly with the decrease in density. The segmentation accuracy of Pointnet embedded with FSDnet increased by 8 %, 4 %, and 11 % respectively under different density conditions; while the accuracy of Pointnet2 embedded with FSDnet

increased by 4 %, 7 %, and 2 %, respectively. This at least shows that FSDnet is very helpful for the network to learn the characteristics of sparse point clouds. The visualization results are as Fig. 12:

#### 4.5. Discussion

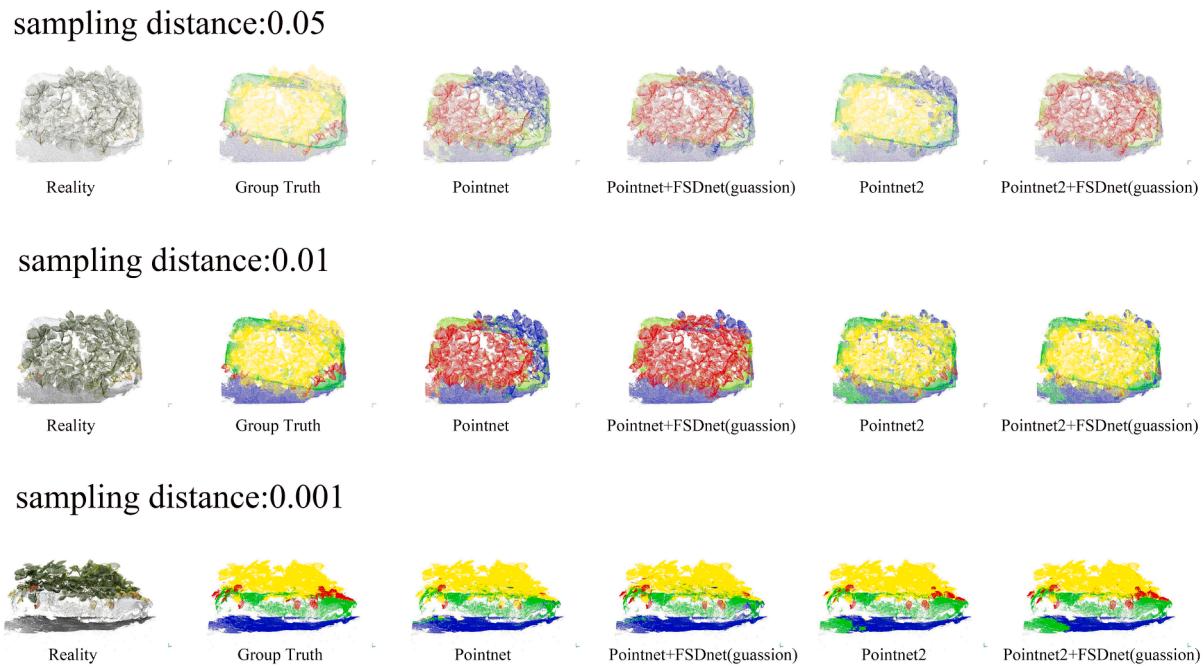
##### 4.5.1. Selection of different density schemes

In this section, we discuss the impact of different density estimation schemes on mIOU, network parameters as Table 6.

On the basis of Pointnet2, we embedded the FSDnet network and added 0.5 M network parameters, which is beneficial to the improvement of mIOU. After that, we adopted three different density estimation schemes, among which the KNN\_max and KNN\_mean density schemes increased the computational traffic by about 0.8G, and the accuracy improved by 4 % and 1 %, respectively. The Gaussian density increases by only 0.32G while the mIOU increases by 3 %, which proves the validity of our gaussian density estimation.

##### 4.5.2. The adaptability of FSDnet on field fruit dataset

We compared the above experiments and found that under the same data set, the lower the point cloud density, the sparser the features, and the worse the model's ability to learn features, which has a great impact on the segmentation of point cloud data captured in agricultural scenarios. However, after the models with medium parameter quantities embed with the FSDnet module, the accuracy of its segmentation has been greatly improved, which shows that FSDnet can assist the main model to learn the sparse features of the point cloud. In addition, comparing different data sets, we found that whether it is in the



**Fig. 12.** Strawberry dataset different density results visualization atlas.

**Table 6**  
Comparison of parameters of different density schemes.

Model\Parameters	mIOU (%)	Flops (G)	Params (M)
Pointnet2	69	2.01	18.6
Pointnet2 + FSDnet(max)	73	2.81	19.1
Pointnet2 + FSDnet (mean)	70	2.80	19.1
Pointnet2 + FSDnet(Gaussian)	72	2.33	19.1

strawberry data set with relatively uneven point cloud, many discrete points, and holes, or in the apple data set with relatively uniform point cloud, the FSDnet module has improved the mIOU of the main network. This shows that FSDnet inputs the density as an auxiliary feature into the network, which is helpful for the network to capture the phenotype of the fruit point cloud.

## 5. Conclusion

In this paper, we propose a novel density-based network module FSDnet for raw point clouds. It provides a feasible solution for the problem of precision degradation caused by the porous and discrete points of image synthesis point cloud. Provide a summary of all the aforementioned experimental results, in agricultural scenarios, FSDnet can improve all models with medium parameter quantities; However, FSDnet has no obvious improvement effect on models with larger parameters. In the comparative experiments of different density schemes, our Gaussian density scheme has better overall performance than other traditional density schemes, and FSDnet also greatly alleviate the reduction in segmentation accuracy caused by the sparseness of fruit features due to density reduction. These reduces FSDnet's requirements on the performance of capture devices, and thus can be adapted for large-scale deployment in orchard.

Although FSDnet has certain value in crop plant phenotype segmentation, automatic field picking and general fields, it is difficult to achieve real-time speed in crop point cloud acquisition and pre-processing based on depth camera, which hinders the further application of FSDnet. At the same time, FSDnet currently has large network parameters. In the future, we will continue to optimize FSDnet, reduce

network parameters, and improve the point cloud acquisition scheme to make it truly applicable to light agricultural picking robot scenarios.

## CRediT authorship contribution statement

**Qinghe Liu:** Conceptualization, Data curation, Funding acquisition, Methodology, Resources, Writing – original draft, Writing – review & editing. **Huijun Yang:** Funding acquisition, Investigation, Project administration, Writing – review & editing. **Junjie Wei:** Writing – review & editing, Conceptualization, Methodology, Software, Validation. **Yuxuan Zhang:** Resources, Project administration, Data curation. **Shuo Yang:** Resources, Project administration, Investigation, Data curation.

## Declaration of competing interest

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

## Data availability

Data will be made available on request.

## Acknowledgment

Research supported by Foundation of Key Research and Development Program of Shaanxi province (2023-YBNY-229, 2024NC-YBXM-19), Undergraduate Training Program for Innovation and entrepreneurship plan (202210712189, S202310712410).

## References

- Boulch, A., Le Saux, B., Audebert, N., 2017. Unstructured point cloud semantic labeling using deep segmentation networks. *3dor@eurographics* 3, 1–8.
- Chang, Angel X., Funkhouser, Thomas, Guibas, Leonidas, Hanrahan, Pat, Huang, Qixing, Li, Zimo, Savarese, Silvio, Savva, Manolis, Song, Shuran, Su, Hao, et al., 2015. Shapenet: an information-rich 3d model repository. *arXiv preprint arXiv: 1512.03012*.
- Chen, Y.i., Xiong, Y., Zhang, B., Zhou, J., Zhang, Q., 2021. 3d point cloud semantic segmentation toward large-scale unstructured agricultural scene classification. *Comput. Electron. Agric.* 190, 106445.

- Dai, A., Chang, A.X., Savva, M., Halber, M., Funkhouser, T., Nießner, M., 2017. Scannet: Richly-annotated 3d reconstructions of indoor scenes. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5828–5839.
- Fangzheng, Hu., Lin, C., Han, J., Peng, J., 2021. Sementing the field of rapeseed from 3d laser point cloud using deep learning. In: 2021 9th International Conference on Agro-Geoinformatics (agro-Geoinformatics), pp. 1–5.
- Jeffrey Hightower, Roy Want, and Gaetano Borriello. Spoton: An indoor 3d location sensing technology based on rf signal strength. 2000.
- Huang, Qiangui, Wang, Weiyue, Neumann, Ulrich, 2018a. Recurrent slice networks for 3d segmentation of point clouds. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2626–2635.
- Huang, Qiangui, Wang, Weiyue, Neumann, Ulrich, 2018b. Recurrent slice networks for 3d segmentation of point clouds. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2626–2635.
- Huang, J., You, S., 2016. Point cloud labeling using 3d convolutional neural network. In: 2016 23rd International Conference on Pattern Recognition (ICPR), pp. 2670–2675.
- Jaremo Lawin, Felix, Danelljan, Martin, Tostberg, Patrik, Bhat, Goutam, Shahbaz Khan, Fahad, Felsberg, Michael, 2017. Deep projective 3d semantic segmentation. In: International Conference on Computer Analysis of Images and Patterns. Springer, pp. 95–107.
- Jayakumari, R., Nidamanuri, R.R., Ramiya, A.M., 2021. Object-level classification of vegetable crops in 3d lidar point cloud using deep learning convolutional neural networks. *Precis. Agric.* 22 (5), 1617–1633.
- Landrieu, L., Boussaha, M., 2019. Point cloud oversegmentation with graph-structured deep metric learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 7440–7449.
- Li, X., Wang, L., Wang, M., Wen, C., Fang, Y., 2020. DANCE-NET: Density-aware convolution networks with context encoding for airborne LiDAR point cloud classification. *ISPRS J. Photogramm. Remote Sens.* 166, 128–139.
- Li, Y., Wen, W., Miao, T., Sheng, Wu., Zetao, Yu., Wang, X., Guo, X., Zhao, C., 2022. Automatic organ-level point cloud segmentation of maize shoots by integrating high-throughput data acquisition and deep learning. *Comput. Electron. Agric.* 193, 106702.
- Mao, J., Wang, X., Li, H., 2019. Interpolated convolutional networks for 3d point cloud understanding. Proceedings of the IEEE/CVF International Conference on Computer Vision.
- Meng, H.-Y., Gao, L., Lai, Y.-K., Manocha, D., 2019. Vv-net: voxel vae net with group convolutions for point cloud segmentation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 8500–8508.
- Ni, P., Zhang, W., Zhu, X., Cao, Q., 2020. Pointnet++ grasping: Learning an end-to-end spatial grasp generation algorithm from sparse point clouds. In: 2020 IEEE International Conference on Robotics and Automation (ICRA), pp. 3619–3625.
- Qi, C.R., Hao, Su., 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.
- Tchapmi, L., Choy, C., Armeni, I., Gwak, JunYoung, Savarese, S., 2017. Segcloud: Semantic segmentation of 3d point clouds. In: 2017 International Conference on 3D Vision (3DV), pp. 537–547.
- Thomas, H., Qi, C.R., Deschaud, J.-E., Marcotegui, B., Goulette, Francois, Guibas, Leonidas J., 2019. Kpconv: flexible and deformable convolution for point clouds. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 6411–6420.
- Wu, W., Qi, Z., Fuxin, L., 2019. Pointconv: deep convolutional networks on 3d point clouds. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 9621–9630.
- Yu, Q., Yang, H., Gao, Y., Ma, X., Chen, G., Wang, X., 2022. Lfpnet: Lightweight network on real point sets for fruit classification and segmentation. *Comput. Electron. Agric.* 194, 106691.
- Zhao, H., et al., 2019b. Pointweb: enhancing local neighborhood features for point cloud processing. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.
- Zhao, H., Jiang, Li, Fu, Chi-Wing, Jia, J., 2019a. Pointweb: Enhancing local neighborhood features for point cloud processing. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 5565–5573.
- Zhiheng, Kang, Ning, Li, 2019. Pyramnet: point cloud pyramid attention network and graph embedding module for classification and segmentation. arXiv preprint arXiv: 1906.03299.