# DIGITAL ORCHARD CONSTRUCTION BASED ON NEURAL RADIANCE FIELD AND GEOREFERENCING TECHNOLOGY
# /
# 基于神经辐射场与地理坐标配准技术的果树三维重建

**Huiyan WANG[1)], Binxiao LIU[2)], Jianhang WANG[1)], Changkun ZHANG[1)], Jinliang GONG[2)], Yanfei ZHANG[1*)]**
[1)]School of Agricultural Engineering and Food Science, Shandong University of Technology, Zibo 255000 / China
[2)]School of Mechanical Engineering, Shandong University of Technology, Zibo 255000 / China
*Tel: +86-18265338441; E-mail: 1392076@sina.com*
*Corresponding author: Yanfei Zhang*
*DOI: https://doi.org/10.35633/inmateh-75-77*

***Keywords:*** *Computer Vision; Fruit Trees; 3D Reconstruction; Neural Radiance Fields; Camera Pose Recovery; Georeferencing*

## ABSTRACT
*This study aims to construct digital fruit trees with high-precision geolocation and high-quality canopy phenotypic details, supporting the development of digital fruit tree technology and the establishment of smart orchards. The Neural Radiance Fields (NeRF) theory was integrated with georeferencing technology. Firstly, multiple ground control points were placed around the tree, and their WGS-84 coordinates were recorded using an RTK surveying instrument. Next, a drone captured multi-view images of the fruit tree, recording the camera poses during the image acquisition. The multi-view fruit tree images undergo ray casting, hierarchical sampling, and high-frequency position encoding before being input into a Multilayer Perceptron (MLP). The MLP was then supervised through volume rendering to obtain a convergent radiance field that reflects the true form of the fruit tree, resulting in the generation of a fruit tree point cloud. Finally, by establishing correspondences between the points in the fruit tree point cloud and the ground control points in the real world, a rigid transformation matrix was computed to convert the point cloud from a local coordinate system to WGS-84 coordinates, yielding a geographically informed digital fruit tree. The experiments demonstrate that the constructed digital fruit tree exhibits excellent phenotypic details and accurately represents multi-scale characteristics. The accuracy of tree morphology indicators, such as tree height, crown length, and width, reached 99.12%, 99.34%, and 99.22%, respectively. Compared to point clouds generated by traditional Structure from Motion-Multi View Stereo (SFM-MVS) methods, the root mean square errors were reduced by 61.24%, 73.48%, and 62.32%, respectively. Additionally, the georeferencing accuracy achieved millimeter-level precision, with registration errors generally below 2 mm. The proposed method can construct digital fruit trees with high geolocation accuracy, detailed phenotypic information, and scale consistency, overcoming key barriers in the development of digital fruit tree technology. It can provide comprehensive data for various production operations in smart orchards.*

## 摘要
*本研究旨在构建具有高水平地理定位精度与高品质叶冠表型细节的数字果树，支持数字果树技术体系与智慧果园建设。将神经辐射场（Neural Radiance Fields，NeRF）理论与地理坐标配准（Georeferencing）技术相结合，以初果期的桃树作为研究对象。首先，在果树周围地面上布设多个地面控制点并通过RTK测量仪记录地面控制点中心位置的WGS-84坐标；其次，使用无人机环绕拍摄果树多视角图像并记录拍摄时相机位姿；然后，将多视角果树图像进行光线投射法分层采样和高频位置编码后输入多层感知机（Multilayer Perceptron，MLP），通过体积渲染（Volume Rendering）监督训练过程以获取收敛且能反映果树真实形态的辐射场并导出果树点云；最后，通过果树点云中与现实世界中地面控制点的对应关系，计算刚性变换矩阵，将果树点云从局部坐标系转换至WGS-84坐标系，得到具有地理信息的数字果树。试验表明，本研究构建的数字果树具有良好的表型细节，可准确表征果树多尺度表型细节。该方法构建的果树点云在树高、冠层长度与宽度等树形指标方面的精度分别达到99.12%、99.34%、99.22%，相较于传统的运动恢复结构-多视图立体匹配（Structure from motion-Multi View Stereo，SFM-MVS）方法构建的果树点云，均方根误差分别减小61.24%、73.48 %、62.32%。同时，其地理坐标配准精度达到毫米级，配准误差普遍小于2mm。该研究提出的方法能构建具有高地理坐标定位精度、高表型细节与高尺度一致性的数字果树，突破了制约数字果树技术体系发展的关键瓶颈，能够为智慧果园的数字化果树表型组学研究、数字化树形管理、数字化生长监测等领域提供关键技术支撑。*

*Huiyan Wang , M.S. Stud.; Binxiao Liu , M.S. Stud.; Jianhang Wang , M.S. Stud.; Changkun Zhang , M.S. Stud.; Jinliang Gong , Professor Dr.; Yanfei Zhang , Professor Dr.*

## INTRODUCTION

In the intelligent orchard production model of all-weather, all-process, and all-space unmanned operations, the digital fruit tree technology system plays a crucial role *(Wu et al., 2021)*. It provides comprehensive spatial perception capabilities for unmanned agricultural machinery, guiding autonomous tasks such as pruning *(Yang et al., 2017)*, thinning *(Zhang et al., 2020)*, and harvesting *(Zhao, 2022)*. Simultaneously, in fields such as digital phenomics research *(Hu et al., 2019)*, digital tree management (*Jiménez-Brenes et al., 2017*), and digital growth monitoring *(Bdulridha et al., 2019)*, the digital fruit tree finds extensive application, holding significant importance for the construction and management of smart orchards *(Zhou et al., 2019)*.

The acquisition of phenotypic information from fruit trees is a critical step in constructing digital fruit trees *(Narvaez et al., 2019)*. The main approaches for obtaining three-dimensional phenotypes of fruit trees are laser scanning systems *(Zhang et al., 2020; Colaço et al., 2017)* and stereo vision systems. Laser scanning provides high accuracy but is costly and lacks color information. In contrast, stereo vision systems, represented by the Structure from Motion-Multi View Stereo *(SfM-MVS)* method, have lower costs and can simultaneously capture both the structure and color information of fruit trees.

Notably, Dongyu Ren et al. achieved the reconstruction of peach tree branches and crowns using the Kinect v2 camera *(Ren et al., 2022)*,

Gatziolis et al. planned various aerial trajectories for SfM-MVS in tree 3D reconstruction *(Gatziolis et al., 2015)*, and Miller et al. reconstructed potted trees using handheld cameras *(Miller et al., 2015)*. However, for fruit tree canopies with complex topological structures and multi-scale high-frequency details, traditional stereo vision systems struggle to accurately capture their three-dimensional phenotypes, posing a critical scientific challenge to the development of digital fruit trees.

Mildenhall et al. introduced the theory of Neural Radiance Fields (NeRF) *(Mildenhall et al., 2021)*, an implicit neural rendering *(Tewari et al., 2022)* for three-dimensional reconstruction. It achieves finer three-dimensional representations of complex phenotypes through sparse input image sets and a Multilayer Perceptron (MLP). With research and improvements to the NeRF theory, there has been a significant leap in both speed and quality *(Barron et al., 2021; Martin-Brualla et al., 2021; Tancik et al., 2023; Müller et al., 2022)*.

The foundation of digital fruit trees lies in obtaining phenotypic information from fruit trees, and further applications rely on accurate geographical information *(Zhang et al., 2013)*. Geographical information supports automated agricultural practices in smart orchards and allows the integration of geographical, environmental, and meteorological data to build a comprehensive digital fruit tree management system. However, the fruit tree point clouds generated by stereo vision systems are limited to local coordinate systems, unable to provide global geographical information. Real-Time Kinematic (RTK) positioning systems, leveraging real-time differential GNSS technology, achieve centimeter-level high-precision geographical coordinate positioning. Therefore, combining the RTK positioning system with fruit tree point clouds allows for georeferencing, integrating digital fruit trees into a global geographical spatial framework. Extensive research has been conducted on the alignment and fusion of point clouds.

*Nistér D., (2004*), extracted feature points from point cloud data for feature point matching after acquiring the data using a stereo camera, thereby efficiently solving the traditional five-point relative pose problem.

*Konolige et al., (2008)* and *Akbarzadeh et al., (2006*), utilized the ICP (Iterative Closest Point) algorithm to align point cloud data from multiple scenes. The ICP algorithm has shown considerable promise in recent years. However, the alignment in this study requires the matching and fusion of two ground point clouds with significant overlap, using image control points. For aligning fruit tree point clouds and georeferenced ground point clouds, this study refers to the four-point fast matching algorithm proposed by *Aiger D et al., (2008)*.

In addressing the challenges of stereo vision 3D reconstruction technology in accurately representing multi-scale complex phenotypic details of fruit trees and the lack of geographical location information in generated fruit tree point clouds, this study combined the Neural Radiance Fields theory with georeferencing technology. Focusing on peach trees, a method for constructing digital fruit trees based on the Neural Radiance Fields theory and georeferencing technology is proposed. The aim is to leverage the NeRF's excellent representation capabilities for high-frequency details and complex topological structures and the RTK's centimeter-level high-precision geographical coordinate positioning ability to build a digital fruit tree with high-level geospatial positioning accuracy and high-quality canopy phenotypic details.

## MATERIALS AND METHODS

The experimental site for this study is the peach orchard at Shandong University of Technology, with the experimental subjects being individual peach trees in the early fruiting stage. Firstly, multiple ground control points were positioned around the fruit tree, and their WGS-84 coordinates were meticulously recorded using an RTK surveying instrument. Subsequently, a drone was employed to capture multi-perspective images of the fruit tree while simultaneously recording the camera poses during image acquisition. Following this, a Neural Radiance Field for the fruit tree was trained using image data augmented with additional pose information, and a point cloud was generated as an output. Lastly, through establishing correspondences between ground control points in the real world and points within the fruit tree point cloud, a rigid transformation matrix was computed. This facilitated the conversion of the fruit tree point cloud from local coordinate systems to the WGS-84 coordinate system, resulting in the generation of a geospatially informed digital fruit tree.

The equipment for fruit tree image acquisition and geographical coordinate measurement is illustrated in Figure 1. The image acquisition device utilized is the DJI "Mavic 2" unmanned aerial vehicle, while the geographical coordinate measurement device consists of the Qianxun SR2 high-precision Real-Time Kinematic (RTK) surveying instrument and ground control points.


**Fig. 1 - Experimental equipment**

*1. Ground control point; 2. UAV; 3 RTK receiver 4. RTK display terminal; 5. RTK positioning rod*

The individual peach tree selected for the experiment measured 2.18 m in height with a crown width of 2.17 m. The data collection was conducted between 4:30 and 4:45 p.m. under cloudy weather conditions, with a light breeze and no direct sunlight. The drone used for image acquisition operated with a focal length of 26 mm and a resolution of 1920 × 1080 pixels. The setup for fruit tree image acquisition and geographical coordinate measurement is shown in Figure 2. Four ground control points (GCPs) were placed around the tree, clearly marked at the base of the trunk, the top of the canopy, and along its edges. Using an RTK surveying instrument, the WGS-84 coordinates of these GCPs were recorded. Following the calibration of the intrinsic parameters of the UAV-mounted camera, the drone captured multi-view images by circling the tree along a spiral trajectory, also illustrated in Figure 2. The flight trajectory was designed based on the experimental protocol outlined by *Tewari et al. (2022)*, featuring a circumferential radius of approximately 3 m, a pitch of 0.5 m, and a rotation speed of around 2 RPM. This process resulted in the collection of 284 images with roughly 50% overlap. Finally, a 3D reconstruction dataset consisting of the fruit tree images and their corresponding camera pose data was compiled, serving as input for the subsequent neural radiance field-based 3D reconstruction of the fruit tree.

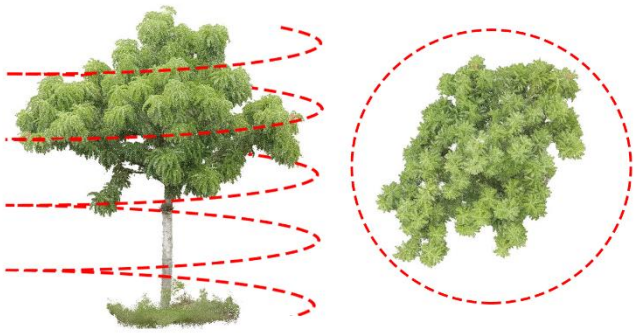


**Fig. 2 - UAV aerial photography path**

The Neural Radiance Fields (NeRF) were trained using two-dimensional fruit tree images and corresponding pose data to transform the two-dimensional images into a three-dimensional spatial structure of the fruit tree. The NeRF utilized an implicit function to record color and volume density information of sampled points in the scene. This function is approximated by a Multilayer Perceptron (MLP) and converges during the training iterations. The experimental hardware configuration for training included: CPU: i9-10850K; GPU: NVIDIA GeForce RTX3090; RAM: 64GB. The software programs used in the point cloud measurement and visualization process were Cloud compare and Metashape.

Initially, camera rays were projected from the camera origin toward specific pixel directions on the image plane using the ray casting method *(Kajyia J et al., 1984)*. Each ray traverses the fruit tree scene to capture visual information along its path. The camera ray D, originating from the camera center and passing through a given pixel, can be mathematically expressed as:

$$\boldsymbol{D}(m,n)=\frac{1}{\left\|K^{-1}\left[\frac{m}{W},\frac{n}{H},1\right]^{T}\right\|_{2}}\begin{bmatrix}\frac{m}{W}\\ \frac{n}{H}\\ 1\end{bmatrix} \tag{1}$$

The variable $(m,n)$ represents the pixel coordinates through which the camera ray passes, with the pixel plane dimensions denoted as $W$ for width and $H$ for height. The camera intrinsic matrix is denoted as $K$.

Subsequently, sampling along the camera ray was conducted, where each sampled point was represented by a vector comprising five parameters: the coordinates of the points sampled along the ray $(x,y,z)$ and the direction of the camera ray $(\theta,\varphi)$. After encoding the sampled points, the vector was input into the Multilayer Perceptron (MLP). The MLP, a fully connected deep neural network, consisted of two parts, as illustrated in Figure 3. The first part of the MLP comprises eight fully connected layers with 256 dimensions each, taking the sampled point coordinates $(x,y,z)$ as input. To address potential issues such as gradient explosion and vanishing gradients in deep neural networks, a residual connection structure *(Rahaman N et al., 2019)* is introduced in the fourth layer of the network. This involves concatenating the output of the fourth layer with the input signal $(x,y,z)$ before inputting it into the fifth layer, thereby breaking network symmetry and enhancing the representational capacity of the MLP. The MLP outputs the volume density $w$ and a 256-dimensional feature vector.

The second part, denoted as cMLP, takes the feature vector outputted by the $w$MLP and a 24-dimensional high-frequency signal representing the direction of the camera ray $f(\theta,\varphi)$ as input. After concatenation, this combined input is passed through a 256-dimensional fully connected layer, and the output from the cMLP layer produces the color $C=(R,G,B)$.

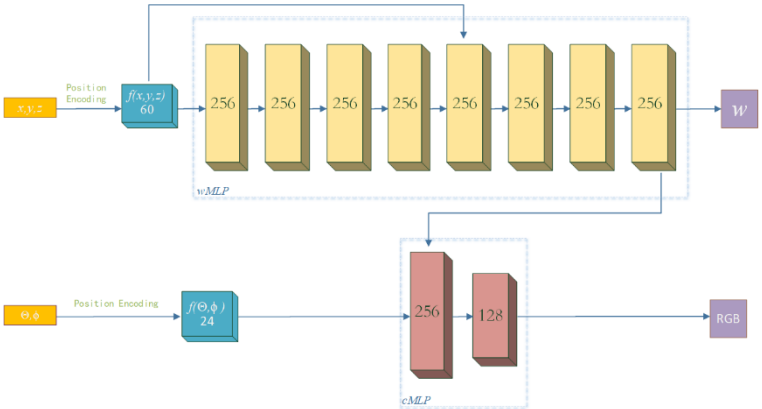



**Fig. 3 - Neural radiance field MLP illustration**

The fruit tree scene was approximated as a neural radiance field describing the volume density and color of all sampled points in the scene through the two components of the Multilayer Perceptron (MLP). The initial neural radiance field generated exhibits a discrete, cloud-like form, lacking accurate representation of the tree's morphology. Therefore, it is necessary to perform volume rendering on the volume density and color of the sampled points outputted by the MLP. Subsequently, the MLP's weights were updated in reverse to facilitate training.

Integral to this process is the direction along the camera ray, where each sampled point is integrated, and the volume density *w* serves as the weighting factor. Based on the volume density *w*, a weighted sum of colors $C$ at each sampled point along the ray is computed. This calculation determines the color of the pixel point traversed by the camera ray. Higher volume density *w* corresponds to lower transparency, thus greater influence of the color $C$ of the sampled point on the pixel color. This computation process is known as volume rendering, and the formula for the calculation is as follows:

$$\begin{cases} C_C(r) = \pi \int_{t_n}^{t_f} T(t) w(r(t)) C(r(t), d) dt \\ T(t) = exp(-\int_{t_n}^{t} w(r(s)) ds) \end{cases} \quad (2)$$

The variable $Cc(r)$ represents the pixel color, $T(t)$denotes the cumulative transmittance coefficient, $w(r(r(t))$ is the volume density of the sampled point at position $r(t)$, $w(r(r(s))$is the volume density of the sampled point at position $r(s)$, $C(r(t))$ represents the color of the sampled point at position $r(t)$, $tn$ is the near point of the view frustum, and $t_f$ is the far point of the view frustum.

Subsequently, to achieve high-quality rendering results with a reduced number of sampled points and simultaneously decrease computational complexity, coarse-fine two-level granularity stratified sampling and rendering were performed along the rays. This involves using the volume density distribution of coarse-grained sampled points as a reference to sample more fine-grained points in regions with higher volume density along the rays.

In the coarse-grained rendering phase, $N_c$ points were uniformly sampled along the camera ray, input into the MLP, and subjected to coarse-grained volume rendering. The resulting coarse-grained rendering pixel color $\hat{C}c(r)$ was obtained. Following this, $N_f$ fine-grained sampled points, based on the volume density distribution of coarse-grained sampled points, are input again into the MLP for fine-grained volume rendering, yielding the fine-grained rendering pixel color $\hat{C}_f(r)$.

$$\begin{cases} \hat{C}_C(r) = \sum_{g=1}^{N_c} w_g C_g \\ \hat{C}_f(r) = \sum_{l=1}^{N_c+N_f} w_l c_l \end{cases} \quad (3)$$

The variables $w_g$ and $w_l$ represent the cumulative volume density of coarse-grained and fine-grained sampled points, respectively, while $C_g$ and $C_l$ denote the colors of coarse-grained and fine-grained sampled points. $N_c$ and $N_f$ are the respective quantities of coarse-grained and fine-grained sampled points.

After computing the coarse-grained rendering pixel color $\hat{C}c(r)$ and the fine-grained rendering pixel color $\hat{C}c(r)$, a comparison was made between each of them and the pixel color $C(r)$ at the pixel $(m,n)$ traversed by the camera ray. This process results in the derivation of the training loss, followed by the computation of the squared sum of the $L$ norm of the loss. Iterating over all pixels for a given viewpoint $V$, the loss function $L$ for that viewpoint is obtained.

$$L = \sum_{r \in V} [\left\| \hat{C}_C(r) - C(r) \right\|_2^2 + \left\| \hat{C}_f(r) - C(r) \right\|_2^2] \quad (4)$$

The variables $\hat{C}c(r)$ and $\hat{C}_f(r)$ represent the pixel colors computed through coarse-grained and fine-grained rendering, respectively, while $C(r)$ is the color of the pixel along the ray direction. $O$ encompasses all pixels in a given view. Subsequently, the network weights of the MLP are updated in reverse based on the loss function, iteratively adjusting the scene to gradually approximate the real morphology of the tree. This process continues until convergence conditions are met, completing the training of the neural radiance field for the tree. The final outcome is the NeRF scene of the tree, which characterized the three-dimensional spatial structure and realistic color information. By mapping the coordinates and color information of each sampled point in the scene to the local coordinate system and recording it, the three-dimensional real-world point cloud of the tree is obtained, thus achieving the three-dimensional reconstruction of the tree.

The fruit tree point cloud obtained from the three-dimensional reconstruction exists in a local coordinate system. To establish a geospatially informed digital representation of the fruit tree, it is imperative to perform georeferencing on the fruit tree point cloud, thereby transforming it from the local coordinate system to the WGS-84 geographic coordinate system. In this study, ground control points (GCP) serve as registration benchmarks. Four ground control points were positioned around the fruit tree. Subsequently, using an RTK, the WGS-84 coordinates at the center of each ground control point were recorded. The simultaneous reconstruction of both the fruit tree and ground control points is illustrated in Figure 4, where the identification numbers of the ground control points are distinctly discernible in the point cloud.

**Fig. 4 - Comparison of ground control point cloud and photo**

Consider the point cloud representing the fruit tree as Point Cloud A. Select the coordinates of four ground control points' centers as matching points, and designate them as the Point Cloud B, with the WGS-84 geographic coordinates recorded by the RTK surveying instrument for these four ground control points as the corresponding matched points. For each matching point and its corresponding matched point, perform centering, and calculate the covariance matrix $H$ using the decentered coordinates of the matching points.

$$\begin{cases} \Delta A_i = A_i - \phi A \\ \Delta B_i = B_i - \phi B \end{cases} \tag{5}$$

$$H = \sum_{i=1}^{4} (\Delta B_i) \cdot (\Delta A_i)^T \tag{6}$$

Where $\phi A$ represents the average coordinates of matching points in Point Cloud A, and $\phi B$ represents the average coordinates of matching points in Point Cloud B. $A_i$ and $B_i$ denote the $i$-th matching points in Point Cloud A and B, respectively. $\triangle A_i$ and $\triangle B_i$ represent the decentered coordinates of the matching points. $H$ stands for the covariance matrix. Subsequently, perform singular value decomposition on $H$ and calculate the scaling factors. The calculation formula is as follows.

$$H = U\Sigma V^{\mathrm{T}} = [u_1 \quad u_2 \quad u_3] \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \end{bmatrix} \begin{bmatrix} v_1^{\mathrm{T}} \\ v_2^{\mathrm{T}} \\ v_3^{\mathrm{T}} \end{bmatrix} \tag{7}$$

$$s = \frac{\sqrt{\sigma_1^{\,2} + \sigma_2^{\,2} + \sigma_3^{\,2}}}{\sigma_1} \tag{8}$$

Where $U$ and $V$ are the left and right singular vector matrices, respectively. $\Sigma$ represents the singular value matrix, and $u_i$ $(i=1,2,3)$ denotes the column vectors of matrix $U$. Similarly, $v_i$ $(i=1,2,3)$ represents the row vectors of matrix $V$, and $\sigma_i$ $(i=1,2,3)$ denotes the singular values, indicating the scaling factors in different directions for the matching points. Through the left and right singular vector matrices and the scaling factor, calculate the rotation matrix $R$ and translation vector $t$ to transform Point Cloud $A$ from the local coordinate system to the WGS-84 coordinate system. These components constitute the rigid transformation matrix that aligns Point Cloud $A$ with Point Cloud $B$.

$$\begin{cases} R = VU^T \\ t = \phi B - s \cdot R \cdot \phi \mathrm{A} \end{cases} \tag{9}$$

$$T = \begin{bmatrix} sR & t \\ 0 & 1 \end{bmatrix} \tag{10}$$

Where $\phi A$ and $\phi B$ represent the average coordinates of matching points in Point Cloud $A$ and $B$, respectively. $U$ and $V$ are the left and right singular vector matrices, $R$ is the rotation matrix, $t$ is the translation vector, $s$ is the scaling factor, and $T$ is the rigid transformation matrix. Applying the rigid transformation matrix $T$ to all points in Point Cloud A aligns Point Cloud A with the coordinate system of Point Cloud B. As a result, each point in the tree point cloud obtains its coordinates in the WGS-84 geographical coordinate system, establishing a digital tree with geographical coordinate information.

## RESULTS AND ANALYSIS

The dataset for the three-dimensional reconstruction of the tree consists of a total of 284 images, with 21 images in the training dataset and 263 in the testing dataset. The training process of the tree NeRF scene spans 30,000 steps, taking 17 minutes and 16 seconds, with a generation rate of 1.14×105 training rays per second. As illustrated in Figure 5, after reaching 15,000 steps, the scene representation stabilizes, and parameters such as learning rate, training loss, distortion loss, RGB loss, among others, gradually converge. The learning rate is a critical component in the optimization of machine learning models. It is employed to regulate the step size of the update process, which is progressively reduced through the implementation of an exponential decay strategy. The training loss functions as the overarching objective function for model optimization, thereby guiding the direction of parameter updates. RGB loss serves as the primary supervisory signal of NeRF, reflecting the model's performance in color reconstruction and geometric rationality. Aberration loss, on the other hand, leads to a more rational volume density distribution, thereby enhancing geometric coherence and rendering quality.

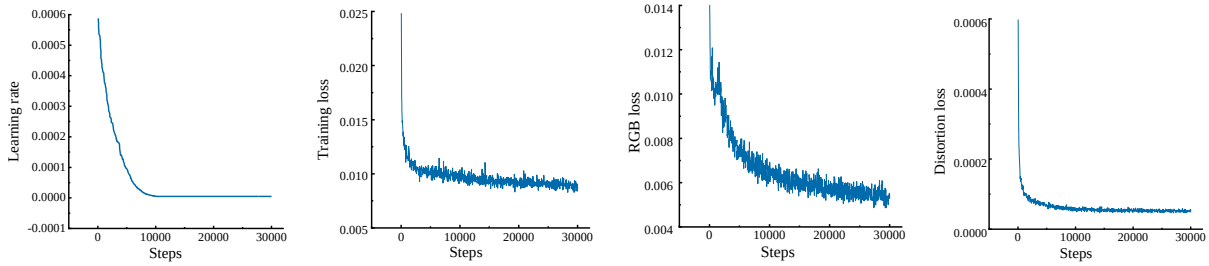



**Fig. 5 - Neural radiance field training parameters**

The reconstructed three-dimensional NeRF scene of the fruit tree is depicted in Figure 6, juxtaposed with corresponding depth maps and photographs of the fruit tree. The fidelity of the reconstructed fruit tree in terms of color and texture is exceptional, closely resembling the photograph of the fruit tree taken at the same pose, thereby achieving a representation effect at the level of real scenes. The contours of the canopy in the NeRF scene depth map are sharp and well-defined, with clear delineation of branches and leaf layers, providing distinct hierarchical depth information. This underscores the commendable three-dimensional representation efficacy of the NeRF scene.

**Fig. 6 - Comparison between fruit tree photo and fruit tree NeRF scene and depth map**

In order to assess the three-dimensional scene representation efficacy of the NeRF method, fruit tree images from the test dataset were compared with corresponding NeRF scene images captured at identical poses. Structural Similarity Index (SSIM) *(Rahaman N et al., 2019)* and Learned Perceptual Image Patch Similarity (LPIPS) *(Zhang R et al., 2018)* were introduced as two image quality assessment metrics to gauge the scene reconstruction capability of the NeRF method. Figure 7 illustrates the evolution of these evaluation metrics during the training process. The LPIPS metric stabilizes within the range of 0.2 to 0.3 after 25,000

steps, with a minimum value of 0.2050. The SSIM metric stabilizes within the range of 0.6 to 0.7 after 25,000 steps, reaching a maximum value of 0.7215.

Both evaluation metrics consistently reside within the high-quality range *(Wang Z et al., 2004)*, indicating the NeRF method's proficient recovery capability for high-frequency detailed scenes.

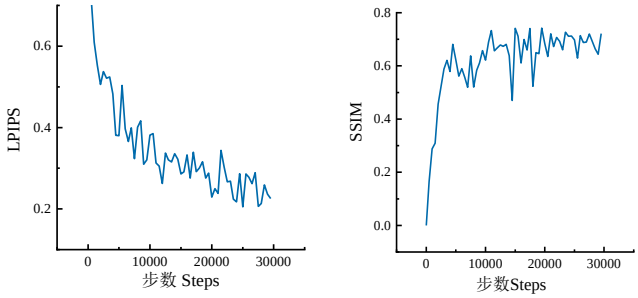



**Fig. 7 - Neural radiance field quality assessment metrics**

Due to the inherent nature of the NeRF scene for fruit trees, which is essentially a representation by an implicit function approximated using Multilayer Perceptrons (MLP), direct editing of the scene is not feasible. Consequently, it becomes imperative to convert the NeRF scene into a three-dimensional point cloud model for subsequent research and applications. The fruit tree point cloud model, derived from the NeRF scene, is depicted in Figure 8, where features such as the fruit tree, ground control points, and yellow markers at the base of the trunk are distinctly reconstructed within the point cloud model.

**Fig. 8 - Comparison between NeRF scene of fruit tree and Point cloud model of fruit tree**

The ground control points play a pivotal role as essential links establishing the connection between the local coordinate system and the WGS-84 coordinate system for the fruit tree point cloud. Four ground control points were deployed for this experiment, denoted as points 2, 13, 20, and 28. The centroids of these four points were utilized as matching points to compute the rigid transformation matrix. Consequently, the ground control points in the fruit tree point cloud model were aligned with their corresponding positions in the WGS-84 coordinate system. In the geospatial coordinate registration process of this experiment, the scaling factor s was determined to be 1.00809. The rigid transformation matrix T applied is expressed as follows:

$$T = \begin{bmatrix} 0.921 & -0.409 & 0.014 & 28.185 \\ 0.409 & 0.919 & -0.058 & -12.497 \\ -0.011 & -0.058 & -1.006 & 58.429 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

To validate the accuracy of geospatial coordinate registration, the right upper corner points of each ground control point and the vertices of the triangular yellow markers at the base of the trunk were selected as test points. An RTK surveying instrument was employed to measure the WGS-84 coordinates of these test points as actual values. Subsequently, the WGS-84 geographical coordinates at the corresponding positions in the fruit tree point cloud after registration were measured as experimental values. A comparison was then conducted between the measured and actual values. The geospatial coordinate registration accuracy, as depicted in Table 1, indicates registration errors within a 2 mm range, attaining accuracy at the millimeter level.

**Table 1**

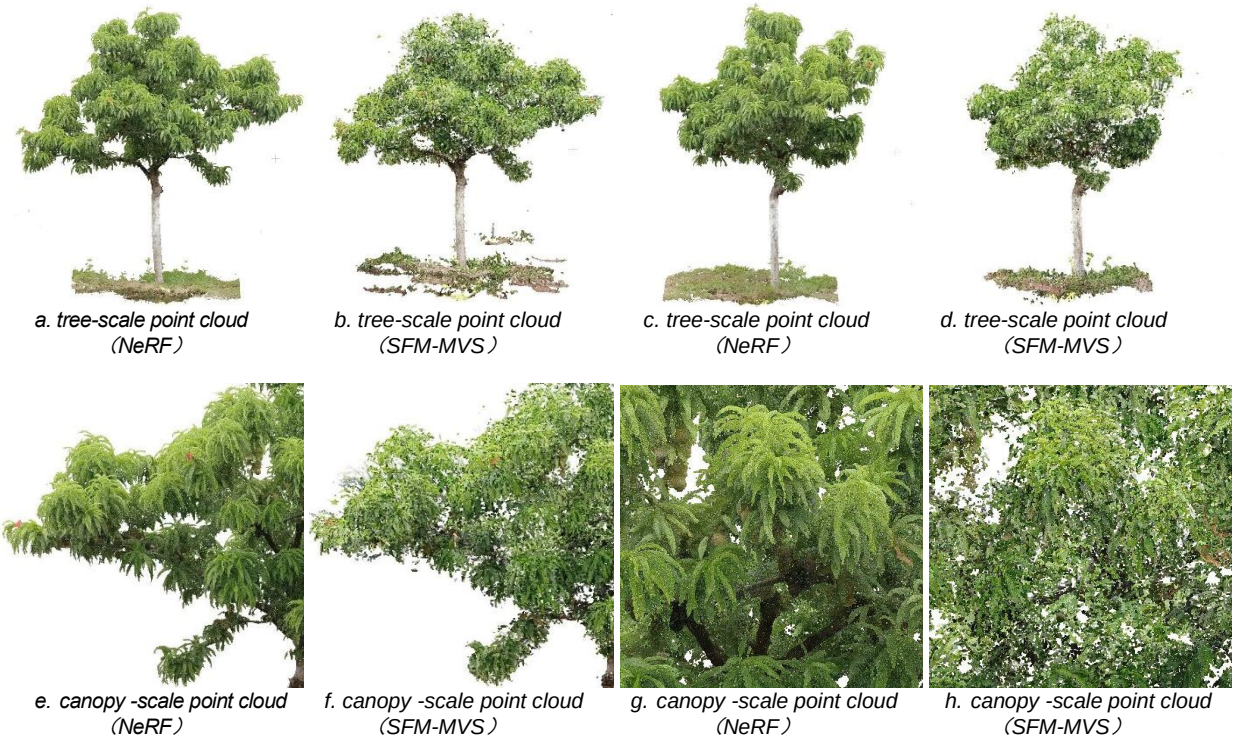
**Georeferencing accuracy evaluation metrics**

| GCPs parameter | X-coordinate | Y-coordinate | Z-coordinate | Registration errors /mm |
|---|---|---|---|---|
| The measured value for GCP 2 | 588758.186722 | 4074069.579384 | 27.214424 | —— |

| GCPs parameter | X-coordinate | Y-coordinate | Z-coordinate | Registration errors /mm |
|---|---|---|---|---|
| The measured value for GCP 13 | 588760.144802 | 4074068.612587 | 27.178391 | —— |
| The measured value for GCP 20 | 588759.746929 | 4074066.781326 | 27.258263 | —— |
| The measured value for GCP 38 | 588757.133163 | 4074067.676079 | 27.354832 | —— |
| Measured value for trunk marker points | 588758.619942 | 4074068.453690 | 27.209908 | —— |
| The tested value for GCP 2 | 588758.171104 | 4074069.567116 | 27.273319 | 1.2001 |
| The tested value for GCP 13 | 588760.156357 | 4074068.628487 | 27.251610 | 1.9655 |
| The tested value for GCP 20 | 588759.755337 | 4074066.795830 | 27.353138 | 1.6764 |
| The tested value for GCP 38 | 588757.148190 | 4074067.669708 | 27.358818 | 1.8644 |
| The tested value for trunk marker points | 588758.608894 | 4074068.453888 | 27.345152 | 1.1049 |

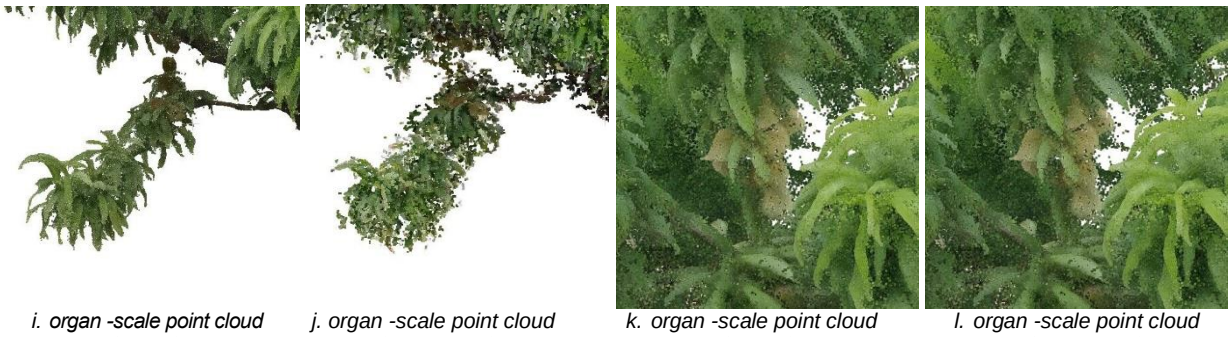
To assess the quality of the NeRF fruit tree point cloud, a set of fruit tree point clouds was reconstructed using the classical stereo vision algorithm SFM-MVS as a control, based on the same dataset. A detailed analysis was conducted to compare the reconstruction effects and reconstruction scale consistency between the two types of fruit tree point clouds.

Regarding the reconstruction effects, as illustrated in Figure 9, a thorough comparison was made at the scales of tree, canopy, and organ for the phenotypic reconstruction of fruit trees using the two methods. At the tree scale, the NeRF method produced a fruit tree point cloud with sharp contours, clear crown textures, and well-defined branch hierarchy. In contrast, the SFM-MVS fruit tree point cloud exhibited blurred contours, abundant noise in the crown, and difficulties in texture recognition. At the canopy scale, the NeRF fruit tree point cloud accurately restored the morphology of branches and leaves. Details such as internal branches, clustered peach tree fruits, and diseased and withered leaves were clearly visible in Figure 9g, highlighting the NeRF method's outstanding recovery capability for high-frequency canopy details. In contrast, the SFM-MVS method at the canopy scale could hardly identify any fruit tree organs. At the organ scale, despite the clustering of immature peach fruits due to non-thinning, the NeRF method could still distinctly reconstruct each fruit, providing a clear, well-organized, and highly recognizable structure. This level of detail is sufficient to guide unmanned agricultural machinery in thinning and harvesting operations. Conversely, the SFM-MVS fruit point cloud at the organ scale exhibited a mosaic-like pattern, rendering it incapable of identifying any details.



*a. tree-scale point cloud (NeRF)* *b. tree-scale point cloud (SFM-MVS)* *c. tree-scale point cloud (NeRF)* *d. tree-scale point cloud (SFM-MVS)*

*e. canopy -scale point cloud (NeRF)* *f. canopy -scale point cloud (SFM-MVS)* *g. canopy -scale point cloud (NeRF)* *h. canopy -scale point cloud (SFM-MVS)*



*i. organ -scale point cloud (NeRF)* *j. organ -scale point cloud (SFM-MVS)* *k. organ -scale point cloud (NeRF)* *l. organ -scale point cloud (SFM-MVS)*

**Fig. 9 - Comparison of NeRF fruit tree point cloud and MVS fruit tree point cloud**

In terms of scale consistency, prior to the experiment, marker points were placed throughout the canopy of the fruit tree, as depicted in Figure 10. Tree height, crown length, and width were measured at these marker points and served as actual values. Additionally, the trunk diameter at 1.3 meters above the ground level was measured as an actual value for breast height.

Within the fruit tree point cloud, the distances were calculated for the corresponding measurement parameters by selecting marker points. Both actual and experimental values were measured 10 times, and the averages were taken as the true values.

**Fig. 10 - Comparison between photo of tagged point and point cloud of tagged point**

As indicated in Table 2, both methods exhibit comparable accuracy levels, with a root mean square error of 0.0047 m, when reconstructing the morphology of objects characterized by regular shape and simplicity, such as breast diameter. They both demonstrate a high-level reconstruction accuracy at the millimeter scale. However, in the reconstruction of the fruit tree canopy, which involves intricate details and complex phenotypic information, the reconstruction accuracy of the NeRF method significantly surpasses that of the SFM-MVS method. The scale consistency accuracy for tree height, crown major axis, and crown minor axis achieved by the NeRF method are 99.12%, 99.34%, and 99.22%, respectively. In comparison to the SFM-MVS method, the root mean square errors are reduced by approximately 61.24%, 73.48%, and 62.32%, respectively.

**Table 2**

**3D Reconstruction scale consistency Evaluation Indexes**

| Parameter | measured value | NeRF tested value | NeRF Root Mean Square Error | NeRF Mean Accuracy | SFM-MVS tested value | SFM-MVS Root Mean Square Error | NeRF Mean Accuracy |
|---|---|---|---|---|---|---|---|
| Tree Height /m | 2.1850 | 2.2030 | 0.0193 | 99.12% | 2.1412 | 0.0495 | 97.73% |
| Canopy Length /m | 2.1760 | 2.1845 | 0.0143 | 99.34% | 2.1157 | 0.0637 | 97.07% |
| Canopy Width /m | 1.6420 | 1.6456 | 0.0128 | 99.22% | 1.6123 | 0.0340 | 97.93% |
| Breast Diameter /m | 0.0680 | 0.0643 | 0.0047 | 93.09% | 0.0645 | 0.0047 | 93.09% |

## CONCLUSIONS

This study, focusing on peach trees in the initial fruiting stage, addressed the scientific challenges associated with the limitations of stereoscopic vision-based 3D reconstruction techniques in representing multiscale complex phenotypic details of fruit trees and the absence of geographic information in generated

fruit tree point clouds. By integrating neural radiative fields with georeferencing technology, a digitally reconstructed fruit tree was successfully created with high-level geographic positioning accuracy and superior leaf canopy phenotypic details, thereby overcoming the technical bottlenecks that have hindered the development of the digital fruit tree technology system.

Experimental results demonstrated that Neural Radiance Fields (NeRF) could capture the intricate topological structure of fruit trees at multiple levels, including tree scale, canopy scale, and organ scale, accurately characterizing morphological features of tree branches, fruits, and even leaves. During the NeRF scene training process, the LPIPS metric reached a minimum value of 0.2050 and stabilized within the range of 0.2 to 0.3, while the SSIM metric achieved a maximum value of 0.7215 and stabilized within the range of 0.6 to 0.7. The scale consistency accuracy of NeRF fruit tree point clouds in tree height, canopy length, and width reached 99.12%, 99.34%, and 99.22%, respectively. Compared to Structure-from-Motion Multi-View Stereo (SFM-MVS) fruit tree point clouds, the root mean square errors were reduced by 61.24%, 73.48%, and 62.32%, respectively. The geographic coordinate registration accuracy of the fruit tree point cloud reached millimeter-level, with registration errors generally less than 2 millimeters.

The digitally reconstructed fruit tree established in this study possesses accurate geographic information and high-resolution phenotypic details at multiple scales. It can provide precise spatial data for unmanned agricultural machinery in the WGS-84 geographic coordinate system, endowing it with global perception capabilities to guide tasks such as pruning, thinning, and harvesting. Additionally, with its high-quality leaf canopy phenotypic details, the digitally reconstructed fruit tree lays a solid technical foundation for research in digital phenomics of fruit trees, digital lighting simulation, digital production management, digital yield estimation, digital growth monitoring, and digital agricultural technology training, thereby holding significant implications for the development of smart orchards.

## ACKNOWLEDGEMENT

The authors thank the anonymous reviewers for their critical comments and suggestions for improving the manuscript. This work was funded by the Key Research and Development Program of Shandong Province (Major Innovative Project in Science and Technology) (2020CXGC010804), Natural Science Foundation of Shandong Province (ZR2021MC026).

## REFERENCES

[1] Akbarzadeh, A., Frahm, J.M., Mordohai, P., Clipp, B., Engels, C., Gallup, D., et.al. (2006). Towards urban 3D reconstruction from video[C] // *Third International Symposium on 3D Data Processing, Visualization, and Transmission (3DPVT'06). IEEE*, pp. 1-8. https://doi.org/10.1109/3DPVT.2006.141. USA.

[2] Aiger, D, Mitra, N.J., Cohen-Or, D. (2008). 4-points congruent sets for robust pairwise surface registration[M] // *ACM SIGGRAPH 2008 papers*.: 1-10. https://doi.org/10.1145/1399504.1360684.

[3] Barron, J.T., Mildenhall, B., Tancik, M., Hedman, P., Martin-Brualla, R., & Srinivasan, P. P. (2021). Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields. American. In *Proceedings of the IEEE/CVF international conference on computer vision*. pp. 5855-5864. USA.

[4] Bdulridha, J., Batuman, O., Ampatzidis, Y. (2019). UAV-based remote sensing technique to detect citrus canker disease utilizing hyper spectral imaging and machine learning. *Remote Sensing*, 11(11): 1373-1395. https://doi.org/10.3390/rs11111373. USA.

[5] Colaço, A. F., Trevisan, R. G., Molin, J. P., Rosell-Polo, J. R., & Escolà, A. (2017). A method to obtain orange crop geometry information using a mobile terrestrial laser scanner and 3D modeling. *Remote Sensing*, 9(8), 763. https://doi.org/10.3390/rs9080763. Brazil.

[6] Gatziolis, D., Lienard, J. F., Vogs, A., & Strigul, N. S. (2015). 3D tree dimensionality assessment using photogrammetry and small unmanned aerial vehicles. *PloS one*, 10(9), e0137765. https://doi.org/10.1371/journal.pone.0137765. USA.

[7] Hu, W., Fu, X., Chen, F., Yang, W. (2019). A Path to Next Generation of Plant Phenomics. *Chinese Bulletin of Botany*, 54(5): 558–568. https://doi.org/10.11983/CBB19141. China.

[8] Jiménez-Brenes, F. M., López-Granados, F., De Castro, A. I., Torres-Sánchez, J., Serrano, N., & Peña, J. M. (2017). Quantifying pruning impacts on olive tree architecture and annual canopy growth by using UAV-based 3D modelling. Spain. *Plant methods*, 13, 1-15. https://doi.org/10.1186/s13007-017-0205-3. Spain.

[9] Kajiya, J. T., Vonherzen, B. P. (1984). Ray tracing volume densities. *ACM SIGGRAPH computer graphics*, 18(3): 165-174. USA.

[10] Konolige K, Agrawal M. FrameSLAM: From bundle adjustment to real-time visual mapping[J]. IEEE Transactions on Robotics, 2008, 24(5): 1066-1077. USA.

[11] Martin-Brualla, R., Radwan, N., Sajjadi, M. S., Barron, J. T., Dosovitskiy, A., & Duckworth, D. (2021). Nerf in the wild: Neural radiance fields for unconstrained photo collections. *In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition.* pp. 7210-7219. USA.

[12] Miller, J., Morgenroth, J., Gomez, C. (2015). 3D modelling of individual Tree using a handheld camera: Accuracy of height, diameter and volume estimates. *Urban Forestry & Urban Greening*, 14(4): 932-940. https://doi.org/10.1016/j.ufug.2015.09.001. New Zealand.

[13] Mildenhall, B., Srinivasan, P. P., Tancik, M., Barron, J. T., Ramamoorthi, R., & Ng, R. (2021). Nerf: Representing scenes as neural radiance fields for view synthesis. Communications of the ACM, 65(1), 99-106. *Nerf in the wild: Neural radiance fields for unconstrained photo collections*. USA.

[14] Müller, T., Evans, A., Schied, C., & Keller, A. (2022). Instant neural graphics primitives with a multiresolution hash encoding. *ACM transactions on graphics (TOG)*, 41(4), 1-15. USA.

[15] Narvaez, F. Y., Reina, G., Torres-Torriti, M., Kantor, G., & Cheein, F. A. (2017). A survey of ranging and imaging techniques for precision agriculture phenotyping. *IEEE/ASME Transactions on Mechatronics*, 22(6), 2428-2439. https://doi.org/10.1109/TMECH.2017.2760866. Chile.

[16] Nistér D. An efficient solution to the five-point relative pose problem[J]. IEEE transactions on pattern analysis and machine intelligence, 2004, 26(6): 756-770. USA.

[17] Rahaman, N., Baratin, A., Arpit, D., Draxler, F., Lin, M., Hamprecht, F., ... & Courville, A. (2019, May). On the spectral bias of neural networks. In *International conference on machine learning*. pp. 5301-5310. PMLR.

[18] Ren, D., Li, X., Lin, T., Xiong, M., Xu, Z., Cui, G. (2022). 3D Reconstruction Method for Fruit Tree Branches Based on Kinect v2 Sensor(基于 Kinect v2 传感器的果树枝干三维重建方法). *Transactions of the Chinese Society for Agricultural Machinery*, 53(S2): 197-203. http://dx.doi.org/10.6041/j.issn.1000-1298.2022.S2.022. China

[19] Tewari, A., Thies, J., Mildenhall, B., Srinivasan, P., Tretschk, E., Yifan, W., & Golyanik, V. (2022, May). Advances in neural rendering. In Computer Graphics Forum. Vol. 41, No. 2, pp. 703-735.

[20] Teunissen, P. J. G., Khodabandeh, A. (2015). Review and principles of PPP-RTK methods. *Journal of Geodesy*, 89.3: 217-240. Australia.

[21] Wang, Z., Bovik, A., Sheikh, H., et al. (2004). Image quality assessment: from error visibility to structural similarity [J]. *IEEE transactions on image processing*, 13(4): 600-612. China.

[22] Wu, S., Wen, W., Wang, C., Du, J., Guo, X. (2021). Research progress of digital fruit trees and its technology system (数字果树及其技术体系研究进展). *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 37(9): 350-360. http://www.tcsae.org/cn/article/doi/10.11975/j.issn.1002-6819.2021.09.039. China.

[23] Tancik, M., Weber, E., Ng, E., Li, R., Yi, B., Wang, T., ... & Kanazawa, A. (2023, July). Nerfstudio: A modular framework for neural radiance field development. In *ACM SIGGRAPH 2023 conference proceedings*. pp. 1-12. USA.

[24] Yang, L., Zhang, D., Xie, R., Luo, J., Wu, C. (2017). Study on Pruning Simulation of Apple Trees at Initial Fruit Stage (初果期苹果树剪枝仿真研究), *Transactions of the Chinese Society for Agricultural Machinery*, 48(S1): 98-102+333. http://dx.doi.org/10.6041/j.issn.1000-1298.2017.S0.016. China.

[25] Zhang, C., Yang, G., Jiang, Y., Xu, B., Li, X., Zhu, Y., Lei, L., Chen, R., Dong, Z., & Yang, H. (2020). Apple Tree Branch Information Extraction from Terrestrial Laser Scanning and Backpack-LiDAR. *Remote Sensing*, 12(21), 3592. https://doi.org/10.3390/rs12213592. China.

[26] Zhao, J. (2022). *System research on robot apple picking path planning based on deep reinforcement learning (基于深度强化学习的机器人苹果采摘路径规划系统研究)* Zibo: Shandong University of Technology. China.

[27] Zhang, G., Li, X., Cheng, S. (2013). Progress and Prospects of 3S Technology Application to Apple Orchard Informatization (3S 技术在苹果园信息化中应用研究的进展与展望). *Geomatics & Spatial Information Technology*, 36(08): 40-44. China.

[28] Zhang, R., Isola, P., Efros, A. (2018). The unreasonable effectiveness of deep features as a perceptual metric. *Proceedings of the IEEE conference on computer vision and pattern recognition*. 586-595. China.

[29] Zhou, G., Qiu, Y., Fan, J. (2018). Research progress and prospect of digital orchard techniques (数字果园研究进展与发展方向). *China Agricultural Informatics*, 30(01):10-16. http://dx.doi.org/10.12105/j.issn.1672-0423.20180102. China.