agronomy

MDPI

Article

RepC-MVSNet: A Reparameterized Self-Supervised 3D Reconstruction Algorithm for Wheat 3D Reconstruction

Hui Liu [1], Cheng Xin [1], Mengzhen Lai [1], Hangfei He [1], Yongzhao Wang [1], Mantao Wang [2] and Jun Li [2,*]

[1] College of Information Engineering, Sichuan Agricultural University, Ya'an 625000, China; 202004026@stu.sicau.edu.cn (H.L.); 202105726@stu.sicau.edu.cn (C.X.); 202105819@stu.sicau.edu.cn (M.L.); 202105766@stu.sicau.edu.cn (H.H.); 202005799@stu.sicau.edu.cn (Y.W.)

[2] Sichuan Key Laboratory of Agricultural Information Engineering, Ya'an 625000, China; wangmantao@sicau.edu.cn

* Correspondence: lijun@sicau.edu.cn; Tel.: +86-130-5659-1398

Abstract: The application of 3D digital models to high-throughput plant phenotypic analysis is a research hotspot nowadays. Traditional methods, such as manual measurement and laser scanning, have high costs, and multi-view, unsupervised reconstruction methods are still blank in the field of crop research. It is challenging to obtain a high-quality 3D crop surface feature composition for 3D reconstruction. In this paper, we propose a wheat point cloud generation and 3D reconstruction method based on SfM and MVS using sequential wheat crop images. Firstly, the camera intrinsics and camera extrinsics of wheat were estimated using a structure-from-motion system with feature maps, which effectively solved the problem of camera point location design. Secondly, we proposed the ReC-MVSNet, which integrates the heavy parametric structure into the point cloud 3D reconstruction network, overcoming the difficulty of capturing complex features via the traditional MVS model. Through experiments, it was shown that this research method achieves non-invasive reconstruction of the 3D phenotypic structure of realistic objects, the accuracy of the proposed model was improved by nearly 43.3%, and the overall value was improved by nearly 14.3%, which provided a new idea for the development of virtual 3D digitization.

Keywords: wheat; point cloud; structure from motion; multi-view stereo; neural rendering

Citation: Liu, H.; Xin, C.; Lai, M.; He, H.; Wang, Y.; Wang, M.; Li, J. RepC-MVSNet: A Reparameterized Self-Supervised 3D Reconstruction Algorithm for Wheat 3D Reconstruction. *Agronomy* **2023**, *13*, 1975. https://doi.org/10.3390/agronomy13081975

Academic Editor: Juncheng Ma

Received: 16 May 2023
Revised: 23 July 2023
Accepted: 24 July 2023
Published: 26 July 2023

1. Introduction

Wheat is a kind of Poaceae plant; as the first food crop grown by humans, it is now the second most produced food crop in the world. Due to the high economic and nutritional value of wheat, its breeding process has been widely studied. Among them, plant phenotypic analysis is a key link to understanding plant gene functions and environmental effects [1], and the analysis of wheat phenotypic parameters is very important for wheat breeding. Currently, data collection in the wheat breeding process is mainly performed manually. These simple but tedious tasks consume a lot of time and energy of researchers and have a high error rate, which affects the research process. So far, the genotypic analysis techniques of wheat and other monocotyledon plants are still relatively scarce, and the cumbersome and inefficient artificial phenotypic analysis methods greatly limit the development of plant gene research. There is an urgent need for a high-precision phenotyping method to make it possible to obtain the true 3D structure of wheat easily and stably.

Plant three-dimensional reconstruction can obtain plant phenotypic data non-destructively [2] and then analyze phenotypic information by combining it with algorithms, such as 3D target detection and point cloud segmentation, which is of great significance for continuous wheat breeding research. The 3D reconstruction of wheat and the establishment and study of the 3D model of wheat will lay a solid foundation for the subsequent study of the 3D dynamic simulation model of the wheat growth and development process. At the same time, the establishment of the 3D visualization model of wheat will also provide

the technical basis for wheat ideal plant type screening and high yield, high efficiency, and lodging resistance design and optimization. The 3D visualization model of wheat morphology constructed by this method has a strong sense of reality. At the same time, this method can also provide a basis for the visualization research of barley, rice, and other crops.

The research on wheat phenotypes has been accumulated for decades, and many scholars at home and abroad have made great contributions in this field. Bo Wang et al. [3] extracted phenotypic features of wheat via supervoxel segmentation of 3D CT images; Y.oda et al. [4] proposed a network of phenotypic data collection for wheat and other crops, which provided a training data set basis for the study of plant phenotypes. S. Lakshmi et al. [5] improved wheat crop yield by using image analysis based on nature-inspired optimization technology to analyze wheat phenotypes. Most of the previous wheat phenotyping studies required a lot of manual work and economic investment. Currently, point cloud technology is rapidly developing to reconstruct data points on the surface of objects into 3D models. Point cloud data is considered to be the closest data set to the original sensor; in addition, point cloud stores three-dimensional information about an object, which has a natural advantage in dimensionality compared to two-dimensional data. Therefore, point cloud data has richer features compared to traditional image data. At the same time, point cloud data avoids the distortion produced by camera photography, which is more convenient for our observation and analysis. The combination of point cloud and wheat breeding can greatly improve computing power and improve the research efficiency of breeding researchers.

In this paper, a new network model of wheat point cloud generation and 3D reconstruction based on MVS and SfM is proposed to provide 3D data for wheat phenotype analysis in response to the problems of existing wheat phenotype research techniques in practical applications. Meanwhile, according to the characteristics of the self-collected data set, the algorithm was improved for reconstructing the complete 3D model of wheat so that it could observe the real wheat phenotypic structure in more detail.

In summary, our contributions are as follows:

(1) This paper takes wheat as the research object and constructs a wheat point cloud generation dataset based on multi-view images to complete phenotypic analysis and 3D reconstruction for wheat and accelerate the research and breeding process.
(2) We propose an integrated framework for non-contact, multi-view 3D reconstruction based on SfM and MVS and introduce various optimization and adjustment strategies to enhance the network performance. The camera parameter matrix information of wheat images was estimated using a structure-from-motion system, which solves the problems of the high cost of data acquisition and easy damage to plant phenotypes caused by previous devices.
(3) We propose the RepC-MVSNet model, which incorporates the RepVGG module decoupled training and inference architecture to enhance the extraction of complex phenotypic features of wheat and can be widely applied to 3D reconstruction tasks of crops.

2. Related Work

2.1. Deep Learning

Wheat is one of the most important food crops in the world and has great relevance to agricultural economic development. In recent years, the organic combination of computer technology and agricultural knowledge has enabled the study of crop phenotypes and morphological structures to step into the visualization stage. With the rapid development of deep learning technology, it has shown great advantages over traditional methods in classification recognition, object detection, and other aspects. Jinya Su et al. [6] used the segmentation algorithm U-Net for yellow rust detection in wheat fields, which can detect pest damage as early as possible; Jianqing Zhao et al. [7] added a micro-scale detection layer based on the YOLOv5 network for field wheat spike detection. Zhiwen

Mi et al. [8] proposed the network framework C-DenseNet to detect stripe rust grading in wheat and embedded CBAM into DenseNet to achieve automated detection of disease classification using an attention mechanism; Bo Gong et al. [9] designed a real-time wheat straw head detection model based on a deep neural network, using spatial pyramidal pooling to improve the learning of wheat features; Gensheng Hu et al. [10] based their studies on deep learning methods for wheat straw detection, constructing generators for self-reversal networks to improve the overall uniformity of wheat and provide data support for wheat phenotype analysis; K.Sandhu Sandhu et al. [11] used a multilayer perceptron and convolutional neural network to predict complex traits in spring wheat breeding, providing a novel idea for crop genetic research. In summary, it can be concluded that the application of deep learning technology in wheat fields can significantly improve the yield and quality of wheat, and this paper mainly explores the reconstruction method of wheat 3D fields under deep learning.

2.2. Three-Dimensional Reconstruction

Most of the current breeding research is limited to two-dimensional image processing of plants, which lacks three-dimensional geometric structure information, making it difficult to extract the three-dimensional morphological data of plants intuitively and effectively. Wheat is a flexible plant with large dynamic variations throughout the growth cycle. The 3D reconstruction of wheat crops can restore real scene information more realistically and also lay a certain foundation for high-throughput plant phenotype analysis.

In recent years, with the development of computer vision and computer graphics, it has become a hot trend to use computer technology to realize the three-dimensional breeding of crops. Daryl M. Kempthorne et al. [12] achieved a virtualized representation of wheat leaves by fitting a parametric linear combination of 3D scan data to the morphological surface of wheat leaves. Hao Zhang et al. [13] used OpenGL to construct a three-dimensional growth system for wheat plants and realized the growth and development process of wheat plants based on the branching structure features of the plants; Nived Chebrolu et al. [14] proposed the reconstruction of non-rigid growth cycle models of plants by constructing invisible Markov chains using plant nodes; A. McElrone et al. [15] used high resolution X-ray computed tomography (HRCT) to collect the 3D structure of plant xylem and thus visualize the 3D structure of the plant vascular system to analyze its structural function; P. Verboven et al. [16] used synchrotron radiation computed laminography (SR-CL) to dissect the three-dimensional structure of leaves to achieve a visual analysis of the mesophyll cell, and the constructed three-dimensional model can provide a corresponding method for calculating the internal material exchange of leaves; S. D. Di Gennaro et al. [17] used a 3D shape method for biomass estimation of the canopy based on RGB images of grapes acquired by a drone to obtain corresponding plant phenotypic deficiencies and showed good performance; Wei Fang et al. [18] proposed a 3D research method based on volume reconstruction of wheat and built a high-throughput phenotypic analysis platform. These methods have basically achieved more accurate 3D reconstruction, but the following problems still exist: (1) The construction of 3D models with CT images has high requirements for equipment, and the scanned 3D images extract features with large sparsity and incompleteness; (2) the 3D reconstruction based on voxel has fundamental limitations in model optimization and detail processing, which cannot exceed the accuracy of one voxel unit; (3) the extraction of plant joint point features requires point cloud data, which is costly to acquire and still requires strong manual intervention.

2.3. Point Cloud Data

In the field of computer vision, the rise of deep learning on point clouds and the continuous improvement of MVS (multiple view stereo) and SfM (structure from motion) algorithms have provided ideas for the 3D reconstruction of wheat. Charles R. Qi et al. pioneered the concept of "point cloud features" [19,20], marking the rise of point cloud deep learning; Haoqiang Fan et al. [21] proposed a single view 3D reconstruction scheme

and obtained usable point cloud data on common objects, such as tables and chairs; Yuhang Yang [22] et al. built a semantic information-based 3D reconstruction method for wheat using an object detection algorithm, text-to-image algorithm, and point cloud reconstruction algorithm, which greatly reduced the reconstruction cost. Yao Yao et al. [23] then proposed the first scheme of using depth to complete a multi-view 3D reconstruction, which solved the problem of the unsatisfactory reconstruction effect under a single view, and then many variant networks on MVSNet appeared [24–31]. Among them, Di chang et al. [30] improved the MVS algorithm by incorporating the idea of a neural radiation field and improved the accuracy of the self-supervised MVS algorithm to a level comparable to the mainstream supervised algorithms. The self-supervised MVS does not require a depth map such as GroundTruth [31] but still requires the acquisition of camera parameters. While the traditional acquisition of camera parameters needs to be performed by calibration, a recent SfM study [32] obtained their camera parameters by inputting a set of 2D images without any calibration work.

3. Materials and Methods

3.1. Materials

In this study, we conducted experiments using public datasets and self-collected datasets. We use the generalization ability of the model to train the MVS network using the DTU dataset, which contains a total of 7 lighting conditions and 128 scenes; each scene corresponds to image sequences, camera parameters, depth maps, and other data under multiple viewpoints, covering most of the scene targets in real life, making the network still have the ability to reconstruct scenes with large changes in lighting and the presence of complex targets more accurately. The format specification of DTU dataset is shown in Figure 1 below.

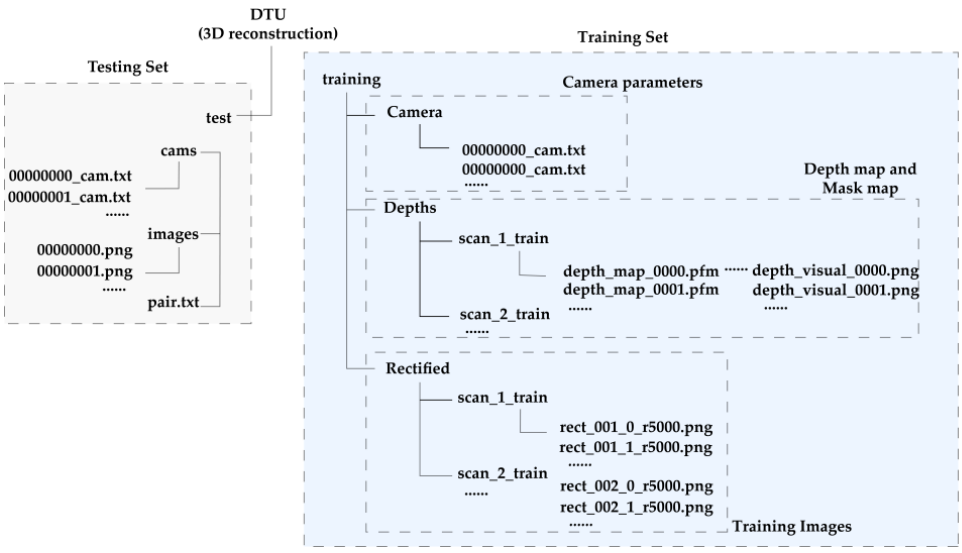

Figure 1. Format specification of DTU dataset.

For the self-collected data set, the data acquisition device was designed based on Raspberry PI for secondary development. The Raspberry PI motionEyeOS video monitoring system was used to shoot and record the growth state of wheat plants in each growth cycle and store the results in the form of pictures and videos. The results were stored in a file or transmitted to a remote server through the network. The camera was equipped with a gimlet built by two micro steering engines to control the shooting angle. When detecting insufficient light, the source of fill light was automatically turned on so as to achieve the best shooting effect. At the same time, the Raspberry PI, as the total controller of the terminal, could send instructions to control the rotating head used to place the plant so that it could rotate periodically or rotate according to the instructions of the operator

so as to realize the multi-directional shooting of wheat plants. In a 3D reconstruction algorithm, one main view and several auxiliary views are often needed. In order to meet the matching requirements of different image pairs, we set the adaptive rotation angle and distance for the rotating head according to the size of the wheat plant so as to cover the effective matching information in the scene as far as possible. A total of 48 wheat images were captured in 5 groups in the experiment with a pixel of 860 × 1200, which is shown in Figure 2 below.

Figure 2. Wheat 3D Reconstruction Dataset (1 group).

3.2. Methods

This study aimed to complete the generation of wheat 3D point clouds. We adopted the conversion of multi-view 2D images to 3D data by combining SfM sparse reconstruction to form sparse point clouds and MVS multi-view 3D reconstruction to form dense point clouds in order to construct the optimal model algorithm.

3.2.1. Structure from Motion System

The first part of this study performs SfM sparse reconstruction of wheat input images where a set of 2D images with common feature points are input to obtain the generated sparse point cloud of wheat $\{Pi\} \in \mathbb{R}^3$, camera intrinsics (Ri, ti) $\in$ SE(3), and camera extrinsics [1]. This method solves the problem of difficult measurement of camera pose parameters for images with different views in the 3D reconstruction process, and the generated camera poses are used as part of the network input in the subsequent MVS reconstruction. We adopted the Pixel-Perfect SFM optimization component, which can be used in any local feature point-based SfM process, as shown in Figure 3, which is mainly divided into two stages of adjustment.

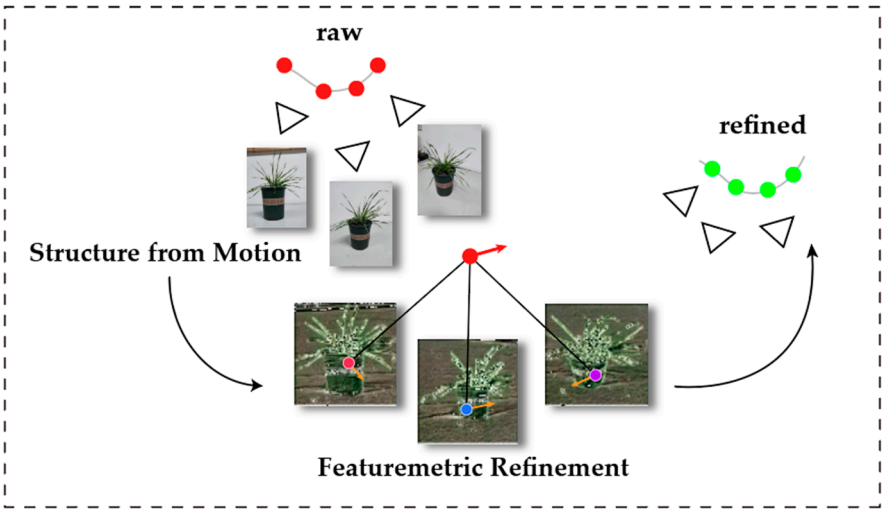

Figure 3. Pixel-Perfect SfM Network architecture.

3.2.2. SfM-Based Network Camera Pose Acquisition for Wheat

The network first uses the S2DNet network to extract a dense map of high-dimensional image features for each of the input wheat images $\{F_i\}$, ensuring broader convergence under challenging conditions.

1. SuperPoint

As a self-supervised end-to-end feature extraction algorithm, SuperPoint uses neural network to extract features and can extract feature points stably. In this study, pixel-perfect SFM was used as the base model for SfM reconstruction, and SuperPoint was used as a feature point extractor to process feature maps.

SuperPoint uses a VGG-like encoder structure to extract features by dimensionality reduction and feeds the features to key point decoder and descriptor decoder, respectively, to obtain the corresponding outputs. In the key point decoder, the corresponding probability magnitude of feature points is calculated for each pixel of the image; in the descriptor decoder, the network obtains the full descriptor via bi-cubic interpolation and L2 norm.

The whole training process of the network (Figure 4) can be divided into three steps: (1) pre-training the key point extraction network, training the network on some datasets of simple objects to obtain a basic decoder; (2) taking the real data without labels as input to obtain pseudo-labels; (3) the real data is transformed geometrically, calculating the true values of key points of the transformed images and feeding this information into SuperPoint for joint training to obtain the final trained SuperPoint. It should be noted that under normal circumstances, the truth values of descriptors cannot be determined, so SuperPoint ensures that the distance of descriptors between the same key points is close enough and the distance of descriptors between different key points is far enough when loss is passed so that the overall training is completed.

2. SuperGlue

In the 3D reconstruction process, the matching process between image pairs is disturbed by redundant information, such as background, due to the environmental factors around the reconstructed objects. Addressing this problem, we used SuperGlue for matching between image pairs, which successfully finds the correct matches between wheat plant features and identifies the incorrect matches between features.

In order to enhance the information perception ability in the 3D scene, the SuperGlue network (Figure 5) introduces the attention mechanism for feature matching. The overall framework of the network can be divided into two parts: the attention-matching layer and the optimal-matching layer. Drawing on the principle of human eyes matching things, the attention-matching layer filters matching points back and forth to enhance the specificity of

feature points. In self-attention, feature points within a single image are aggregated and matched to extract feature points with specificity, while in cross-attention, feature matching is performed between pairs of images. In the optimal matching layer, a soft assignment matrix is constructed to obtain the matching score and matching confidence for each feature point in the two images. For each feature point, the highest score is selected from all scores as the final matching result and then filtered based on the confidence level to remove the error points.

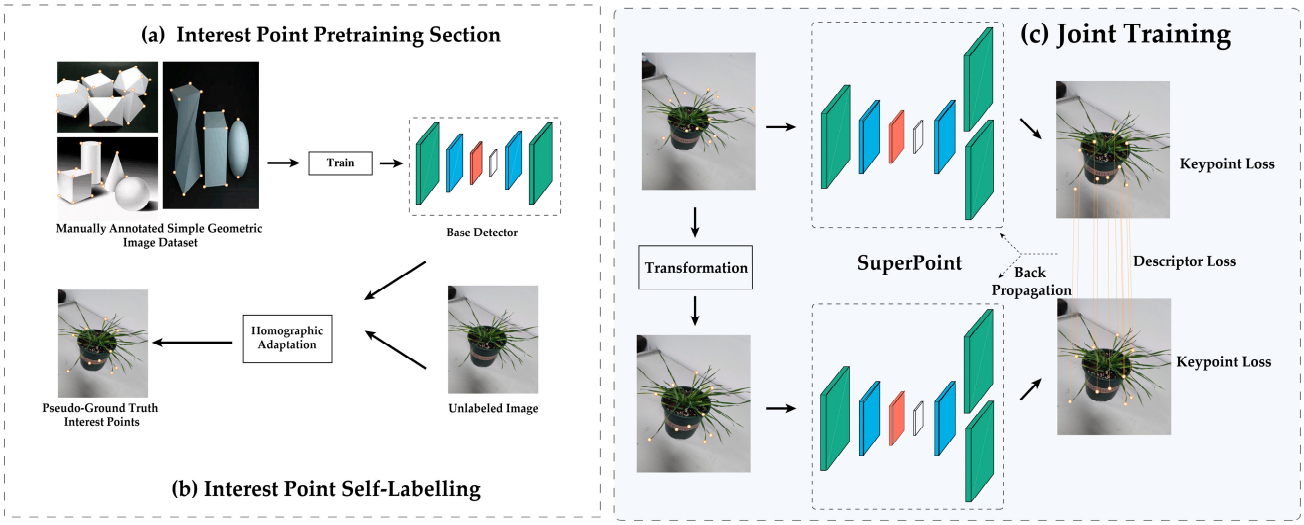

Figure 4. SuperPoint self-supervised training strategy.

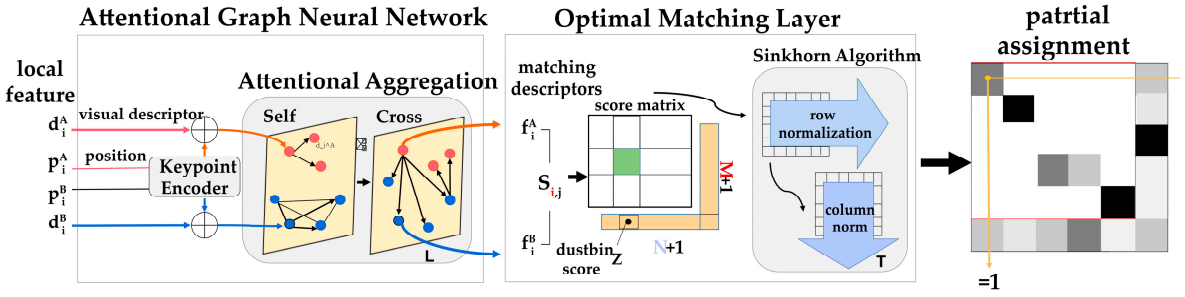

Figure 5. SuperGlue Network structure.

3. Featurematric Key Point Adjustment

After the feature matching process, we can obtain the initial track, triangulate the track to obtain 3D points, adjust the position of each track corresponding feature point in the image, reduce the detection noise by jointly optimizing the graph, and construct the featurematric key point adjustment (FKA) error quickly. Removal of mismatched points:

$$E^{j}_{\mathrm{FKA}} = \sum_{(u,v)\in\mathcal{T}(j)} w_{uv} \| F_{i(u)}[p_u] - F_{k(v)}[p_v] \|_{\gamma} \quad (1)$$

where $\mathcal{T}_j$ refers to a track, and u, v is the matched pair of points on the two images; w_{uv} denotes the confidence of association of the feature point u to feature point v, which can be expressed by the cosine distance of the feature descriptor $d_u^T d_v$; $F_{i(u)}[p_u]$ refers to the features of points on the image. $\|\cdot\|_{\gamma}$ is a parametric norm method with high robustness, which limits the adjustment range of the adjustment point to K.

The refined key points are forward propagated to the standard SfM pipeline to estimate the wheat 3D reconstruction, and a reference descriptor f is extracted for each 3D point. The BA optimization residuals in the reconstruction process are changed from reprojection error to featuremetric error, and the key points in the query image are refined by minimizing the

featuremetric error of the reference to achieve the optimized effect of wheat 3D structure and camera pose; the descriptor optimization BA is calculated as follows:

$$\mu^j = \underset{\mu \in \mathbb{R}^D}{\operatorname{argmin}} \sum\nolimits_{\mathrm{f} \in \{\mathrm{f}_u^j\}} \| \mathrm{f} - \mu \|_\gamma \tag{2}$$

$$f^j = \underset{f \in \{f_u^j\}}{\operatorname{argmin}} \left|\left| \mu^j - f \right|\right| \tag{3}$$

$$E_{\mathrm{FBA}} = \sum\nolimits_j \sum\nolimits_{(i,u) \in \mathcal{T}(j)} \| F_i\left[\Pi\left(R_i P_j + t_i, C_i\right)\right] - f^j \|_\gamma \tag{4}$$

where $F_i\left[\pi\left(R_i P_j + t_i, C_i\right)\right]$ is the feature map interpolation at projected location, and f^j is the reference descriptor.

Our SfM reconstruction method improves the sparse motion structure through depth features in multiple views, uses deep learning to extract depth features, and carries out quadratic feature matching. The use of locally dense information is much more accurate than geometric optimization, which greatly improves the accuracy of 3D points and camera posture, generates subpixel accurate reconstruction, and can process thousands of images with very little overhead. Using a set of unstructured images captured from various perspectives, the camera pose is restored, and the sparse 3D point cloud of wheat scene is reconstructed, which effectively solves the problem of camera point design.

3.2.3. MVS

The current image-based 3D reconstruction methods are mainly divided into 3 major categories: monocular depth estimation, binocular stereo matching, and multi-view 3D reconstruction. Among them, the depth map of monocular depth estimation mostly lacks the geometric consistency constraint of multi-views, and the reconstructed 3D geometric effect is poor; the method of binocular stereo matching is limited in application, and the estimated depth value is highly dependent on the focal length and focus of binocular cameras, and for the reconstruction task of large scenes, how to obtain a large baseline distance to a certain extent limits the effect enhancement of 3D reconstruction. In contrast, multi-view stereo matching uses the similarity of multiple views for depth map prediction, which achieves the effect of low data acquisition cost and wide application range.

Under the actual crop growth and cultivation conditions, due to the large variation in the phenotypic growth of the wheat to be monitored and analyzed, the complicated and interlaced branches and leaves can easily lead to serious occlusion situations. How to use 3D point cloud data to achieve phenotypic analysis of wheat plants at the 3D level is one of the key methods to get rid of the previous problems, such as large spatial orientation limitations and incomplete detection of detailed textures for wheat phenotypic monitoring. This paper aims to recover the 3D point cloud of wheat in real scenes from multi-view wheat plant images and corresponding calibrated camera positional parameters and to learn the 3D geometric features of real scenes via unsupervised methods.

3.2.4. MVS-Based Network for 3D Reconstruction

In this paper, RepC-MVSNet algorithm is proposed to realize point cloud generation and 3D reconstruction of wheat. RepC-MVSNet network is built based on the RC-MVSNet algorithm model, which is mainly divided into four core parts. In the first part, the unsupervised trunk CasMVSNet [27] is used to predict the initial depth map of the image, and the obtained initial depth map is used as the depth prior to the subsequent network. In the second part, implicit neural volume construction was carried out by means of neural rendering, and each image was constructed as a source image to synthesize the reference perspective. The third part is the rendering of the image after reference perspective synthesis loss supervision; the fourth part is to monitor the depth consistency loss of the rendered depth. The algorithm architecture is shown in Figure 6 below:

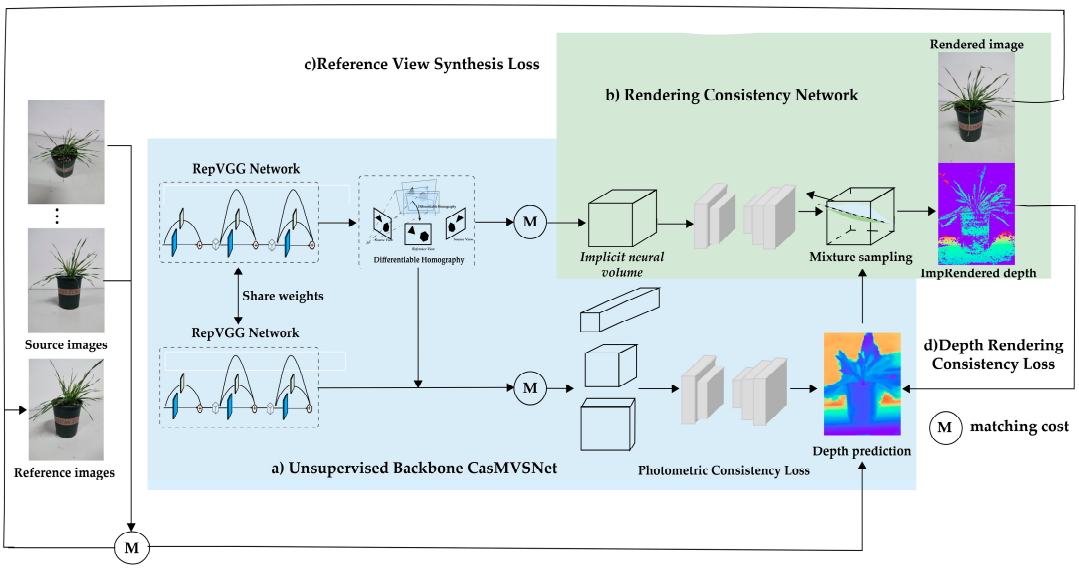

Figure 6. RepC-MVSNet 3D reconstruction algorithm structure diagram.

Compared to most multi-view stereo (MVS) models, RC-MVSNet's innovation takes into account the photometric loss error caused by photometric inconsistency, introduces rendering consistency to replace the photometric consistency of the original unsupervised method, and introduces the reference of neural body rendering. The synthetic loss of view angle is introduced to construct the RGB supervised signal, which eliminates the error of photometric loss caused by different orientation views and enhances the generalization ability of the model. At the same time, wheat as a grass plant has complex plant morphology and many tillers, and the cross-obscuring phenomenon of each branch and leaf causes great interference to the reconstruction process. In the occlusion area, the geometric texture information of the realistic reconstruction target is lost and cannot be matched with that of another viewpoint for the area, which leads to unreliability in matching and reconstruction. Gaussian-uniform mixture sampling can be used to solve this problem.

(1) RepC-MVSNet Network

As an unsupervised ulti-views stereos scheme, first input wheat images $\{I_i\}_{i=1}^{n}$ from multiple perspectives. In the processing of N views, select 1 view as the reference I_1 view, the remaining N-1 view as the source view $\{I_i\}_{i=2}^{n}$ as input. The CasMVSNet backbone network is used to process, and a coarse-to-fine structure is introduced for predicting the depth map instead of using the full-resolution depth estimation method. In this paper, the proposed improvement is to replace part of the original ordinary convolutional neural network for feature extraction with a RepVGG network for the reparameterization process. The RepVGG network uses a multi-branch model with residual-assisted connectivity similar to ResNet in the training phase and a single-way model of VGG type in the inference phase, as shown in Figure 7. In the structural reparameterization process, the merging of the 3×3 convolutional layer and the BN layer is performed first, where the 1×1 convolutional layer is converted into a 3×3 convolutional layer and the residual structure of the BN layer is converted into a 3×3 convolutional layer to form a 3-way 3×3 convolutional layer where the combined convolutional and BN layers are calculated as follows:

$$Conv(x) = W(x) + b \tag{5}$$

$$BN(x) = r\frac{(x - mean)}{\sqrt{var}} + \beta \tag{6}$$

$$BN(Conv(x)) = r\frac{(W(x) + b - mean)}{\sqrt{var}} + \beta = \frac{r \cdot W(x)}{\sqrt{var}} + (\frac{r \cdot (b - mean)}{\sqrt{var}} + \beta) \tag{7}$$

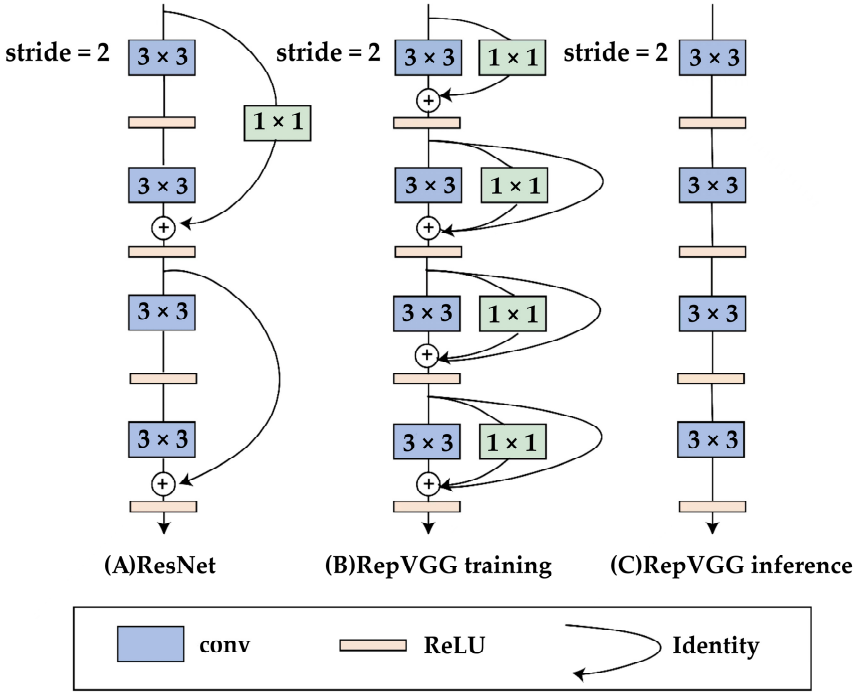

Figure 7. RepVGG algorithm architecture.

The obtained 3-way 3 × 3 convolution uses the additivity operation of convolution to perform multi-branch fusion to form a single 3-way 3 × 3 convolution layer to finally obtain the output result with the same resolution. This operation achieves the advantages of high performance when training multi-branch models and is fast and memory-saving when inferring single-way models.

The input image generates feature map F_i via RepVGG network. The feature map F_i is mapped to the reference camera frustum to generate feature volume $\{V_i\}_{i=1}^{n}$ through differentiable homography.

All the feature volumes obtained from the N times process of differentiable homography are stitched together to calculate the feature variance to construct the cost matrix C, which contains all possible depth ranges, and the matrix C is calculated as follows:

$$C = Var(V_1, V_2, \ldots, V_N) = \frac{1}{N}\sum_{i=1}^{N}(V_i - \overline{V_i})^2 \quad (8)$$

The cost matrix C is optimized by 3D U-Net and obtains the initial depth graph D (Figure 8) under the reference view, which provides a depth prior to the subsequent network processing. After obtaining the depth map, we use the depth map and camera intrinsics and camera extrinsics to realize the micro-reprojection through differentiable homography and bilinear interpolation and project the image of source view to reference view, and we calculate the corresponding valid mask during homography warping through the projection mapping relationship. The difference between the reprojected image and the reference image is calculated in the valid region, and then the error is obtained.

3.2.5. Point Cloud 3D Reconstruction Scheme

After obtaining the dense point cloud, we need to transform the point cloud into a shape model. The most important process of 3D point cloud reconstruction is the normal vector feature calculation, and its mathematical principle is as follows:

$$\begin{cases} P\left(\vec{n}, d\right) = \underbrace{\arg min}_{(\vec{n},d)} \sum_{i=1}^{k} \left(\vec{n} \cdot p_i - d\right)^2 \\ M = \frac{1}{k} \sum_{i=1}^{k} (p_i - p_0)(p_i - p_0)^T \end{cases} \tag{9}$$

where Formula (1) represents the local plane P in the least square sense of scanning point p and its K nearest neighboring points. $\vec{n}$ is the normal vector of the plane P, and d is the distance from P to the origin of the coordinates; Formula (2) represents a covariance matrix that can be solved by PCA on it to obtain the eigenvalue $\vec{n}$.

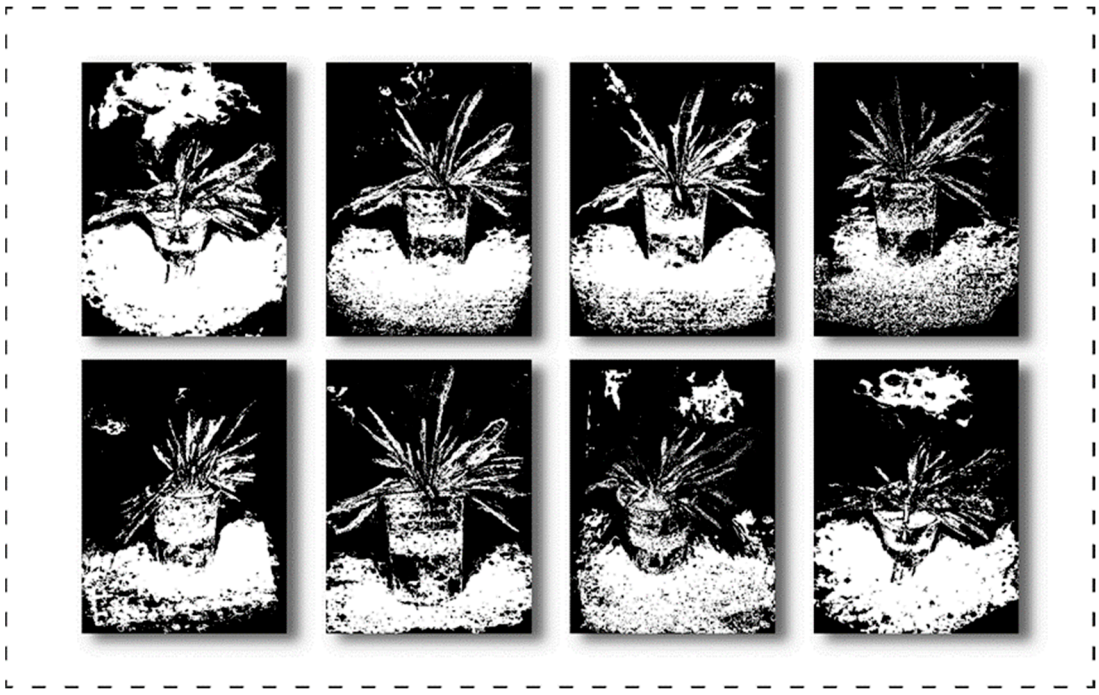

Figure 8. Wheat Mask images.

4. Experiments

4.1. Experimental Details and Evaluation Indicators

The experimental environments were Ubuntu (Canonical Ltd., London, UK), CUDA 11.3 (NVIDIA Corporation, Santa Clara, CA, USA), Python 3.8 (Python Software Foundation, New Castle, DE, USA), PyTorch 1.10.0 (Facebook Artificial Intelligence Institute, New York, NY, USA). The hardware configuration used was NVIDIA GeForce RTX 3090 GPU, 15 vCPU Intel (R) Xeon (R) Platinum 8338C CPU @ 2.60 GHz.

To verify the accurate validity of this experiment, we evaluated the SfM reconstruction task and the MVS reconstruction task, respectively, and the following are the definitions of our evaluation metrics:

4.1.1. SfM Evaluation Indicators

In this paper, we use RANSAC [33] fitting data points, 3D points, average track length, and minimum reprojection error to evaluate the SfM system.

In the process of performing absolute camera pose estimation, PnP and random sample consensus (RANSAC) algorithms are used to extract and optimize the valid sample data and eliminate the matching pairs that do not satisfy the conditions. For the initial sample data, the minimum variance estimation algorithm is used to calculate the final model parameters. When the deviation is <threshold, the sample points are classified as inliers, which is used to describe the final model; when the deviation is >threshold, the sample

points are classified as outliers, which is generally generated by the wrong estimation calculation and cannot meet the needs of the model.

Three-dimensional points refer to the number of three-dimensional points formed after triangulation. During triangulation, 2D alignment images are continuously added to optimize the existing 3D point set through 2D–3D correspondence, which finally produces the final 3D spatial points. This index can indirectly reflect the matching accuracy between the camera's poses and 2D points. When the number of feature extraction points is within the normal threshold period, the more 3D reconstruction points, the higher the matching accuracy.

The average track length refers to the average matchable length of key points in multiple views; the larger the index is, the more key points can be successfully matched in more images, and the more reliable the recovered structure is, but the corresponding will bring ambiguous information, resulting in larger minimum reprojection errors.

The minimum reprojection error is the difference between the observed projection point of the source image $\{p_1, p_2, \ldots\}$ and the projection points $\{\hat{p}_1, \hat{p}_2, \ldots\}$ obtained from the spatial 3D points according to the projection mapping, and the smaller this metric is, the more accurate the recovered structure is. Suppose the spatial point is P_i, the coordinate of the observed projection point of the source image is u_i, and the depth is s_i, the camera intrinsics is K, the camera extrinsics are $\{R,t\}$, and the minimum reprojection error is ξ^*, which is calculated as shown in the following equation:

$$\xi^* = arg\min_{\xi} \frac{1}{2} \sum_{i=1}^{n} \left|\left| u_i - \frac{1}{s_i} KRtP_i \right|\right|_2^2 \tag{10}$$

4.1.2. MVS Evaluation Indicators

In the MVS reconstruction method, the evaluation metrics are accuracy and completeness based on the existing field point cloud. In the DTU dataset, accuracy (Acc.) is defined as the distance between the reconstruction result and the groundturth, which describes the accuracy of the projection result when projecting from the groundturth to the 2D image and is essentially the error distance on the 2D image. Completeness (Comp.) is defined as the distance between the true value and the reconstruction result, which overall is the average of accuracy and completeness.

4.2. Experimental Results

We use several different feature extractions and matchings to select the optimal combination for our SfM task in the first stage and to analyze for the MVS reconstruction task in the second stage. The accuracy of feature extraction is evaluated firstly in Section 4.2.1; the effect of different feature extraction, matching, and optimal adjustment strategies is analyzed via ablation experiments in Section 4.2.2, the model reconstruction error of different MVS algorithms is analyzed in Section 4.2.3, the time consumption of the SfM-MVS algorithm system is analyzed in Section 4.2.4, the depth estimation analysis is performed in Section 4.2.5, and finally, the final wheat point cloud reconstruction results are presented in Section 4.2.6.

4.2.1. Feature Extraction

We used the hand-made local feature SIFT [34] and the deep learning-based SuperPoint [35], D2Net [36], and R2D2 [37] algorithms for comparison in order to find the optimal strategy. According to Figure 9 and Table 1, the number of inlier feature points extracted by SIFT and SuperPoint are 548 and 338, respectively, but the deep learning method SuperPoint, compared with the traditional local feature method SIFT, extracts feature points that match the real detailed features of wheat more closely, and the percentage of inlier in the total data points is 96%, which is 10% higher than that of SIFT. The inliers and outliers extracted by D2Net are much more than the actual target feature body of wheat, and basically no features are extracted effectively, and the inliers and outliers extracted by

R2D2 have a large number of outliers, which belongs to the extraction of a large number of background features incorrectly as wheat targets, and this part of features is easy to introduce noise and has instability.

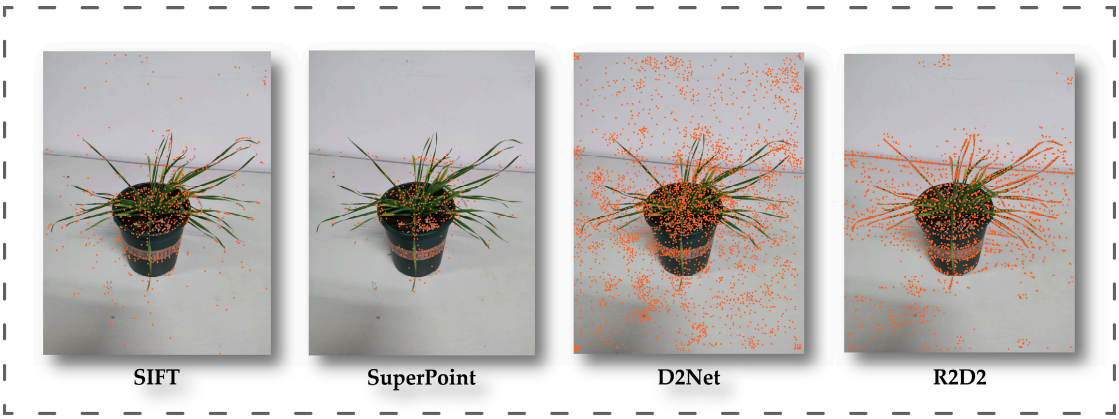

Figure 9. Keypoint extraction results of different algorithms.

Table 1. Comparison of different feature extraction algorithms.

Method	Evaluation Metrics	
	Numbers of Inliers	Numbers of Outliers
SIFT	548	89
SuperPoint	338	13
R2D2	1244	898
D2Net	2309	1796

4.2.2. Ablation Study of SfM

In this paper, we tested two feature extraction algorithms, SuperPoint and D2Net, and two feature matching algorithms, SuperGlue and NN, and conducted ablation experiments by combining two optimization strategies, feature key point and bundle adjustments (FKA and FBA), when FKA and FBA are not used by default, using BA optimization. The experimental results are shown in the following Table 2:

Table 2. Ablation experimental results of SfM system.

Number	Feature Extraction		Feature Matching		Adjustment Strategies		3D Point	Average Track Length	Minimum Reprojection Error
	SuperPoint	D2Net	SuperGlue	NN	FKA	FBA			
01	√		√				4521	7.2245	1.2785
02	√		√		√		4564	7.1422	1.1892
03	√		√			√	4545	7.2033	1.2734
04	√		√		√	√	4570	7.1271	1.1791
05	√			√			3834	6.3560	1.1614
06	√			√	√		3804	6.4189	1.0632
07	√			√		√	3803	6.3986	1.1693
08	√			√	√	√	3804	6.4180	1.0619
09		√		√			22,011	4.5455	1.4426
10		√		√	√		21,996	4.5247	1.3865
11		√		√		√	21,981	4.5526	1.4414
12		√		√	√	√	22,088	4.5186	1.3947

We validate our algorithm design via extensive ablation experiments in the above table, and the experimental results show that (1) SuperPoint + SuperGlue is the best feature

extraction and matching model, the average track length of SuperPoint is 49.6% more than that of theD2Net algorithm, the minimum reprojection error is reduced by 17.2%, the average number of reconstructed 3D points is 19.4% higher than that of the NN algorithm, the average track length is 31.2% more than that of the NN algorithm, and the minimum reprojection error is reduced by 2.7%. SuperPoint + SuperGlue shows good performance; (2) the feature optimization method proposed based on pixel-perfect SfM can effectively refine the overall 3D SfM model when extracted feature points are suitable, KFA can definitely increase the number of correlated matches, which improves the probability of bit pose estimation in the 3D point matching process, and the FBA fine-tuning on this basis reduces the reprojection error by 8%, which can further reduce the feature metric error; (3) FKA is adjusted by topological center selection connecting feature points, and FBA feature center optimization is adjusted when the feature extracted feature points are unstable, and the introduction of the additional noise points make FKA and FBA a limited improvement to the model, while the FKA operation will reject the correct but noisy matching and mislead the global optimization (see 9–12 groups of experimental results); (4) according to the comparison experiment between SuperPoint and D2Net, the 256-dimensional extraction descriptor can better match the real geometric texture for the feature point extraction of wheat, and the accuracy of features with fewer dimensions decreases obviously. In the extraction of higher dimensions, there will be too much dense overlap of descriptors, and the reconstructed 3D point cloud value is far more than the normal value. The introduction of redundant noise information will increase the computing requirements correspondingly; and (5) through the ablation experiments, we conclude that the SfM system of SuperPoint + SuperGlue + FKA + FBA is more compatible with the wheat 3D reconstruction task and produces more accurate visual localization.

In this paper, we also visualize the computational results of the above experiments. As shown in Figure 10, the distribution of feature points of the image is shown on the left side, the orange points $\{P\}$ represent the feature points of the reference image, and the green points $\{P'\}$ represent the feature points mapped from 3D to 2D by the transformation matrix. The orange points and green points basically match, proving the effectiveness of the matching algorithm. On the right side is the sparse point cloud $\{p\}$ triangulated according to the matching result. The red points are the 3D points before optimization, and the green points are the 3D points after optimization, the reprojection error of the latter is smaller than the former, and the completeness and accuracy are higher. The visualization results show that the SuperPoint + SuperGlue and SuperPoint + NN groups have relatively better results with the feature points matching the real scene more closely, the restored preliminary 3D geometric structure is clearer and more explicit, and the fine-tuning is carried out in a relatively small range, while D2Net + NN has many unreliable feature points interfering with the matching due to the excessive number of extracted features. The matched 3D points are also diffused, and the feature details are basically not restored.

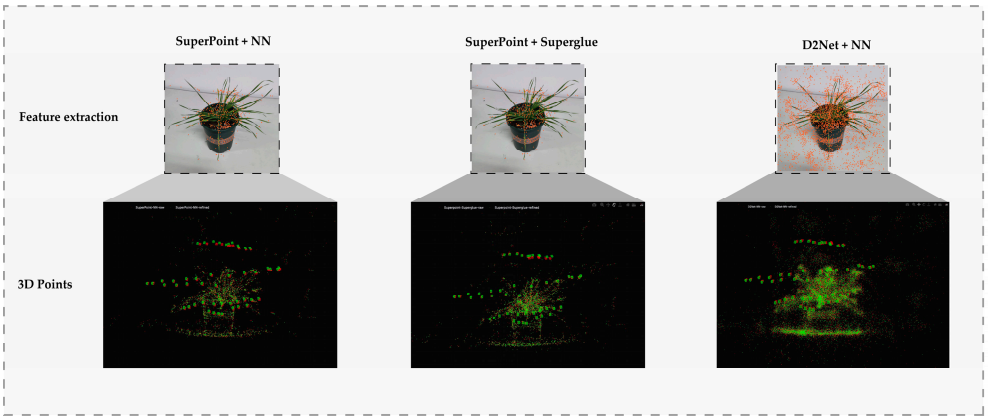

Figure 10. Comparison of feature point extraction (**left**) and structure from motion (**right**) results of SfM system.

4.2.3. MVS Model Evaluation

Due to the lack of supervised information, this paper cannot directly measure our proposed training strategy using a real wheat dataset, but it can indirectly reflect its effectiveness by testing on a benchmark dataset. We conducted experiments on JDACS, RC-MVSNet, and RepC-MVSNet (ours) point cloud reconstruction algorithms by using them on scan1 of the DTU dataset, and the results are shown in the following Table 3. The experimental results show that by adding Repvgg to enhance feature extraction, the accuracy of the proposed model is 0.259, and the completeness is 0.312. Compared to the existing model, our proposed model outperforms all other models we tested, the proposed model is improved by nearly 43.3%, and the overall value is improved by nearly 14.3%, reaching the level of practical application.

Table 3. Evaluation of the impact of each loss of the algorithm on training.

Method	Evaluation Metrics (DTU Dataset)		
	Acc.	Comp.	Overall
JDACS	0.419	0.257	0.338
RC-MVSNet	0.368	0.284	0.326
RepC-MVSNet (ours)	0.259	0.312	0.285

4.2.4. Time Consumption

For the 3D reconstruction process of wheat, we divided the required reconstruction time into two parts, which are the SfM reconstruction to obtain the camera parameters calculation time and the MVS reconstruction to obtain the 3D point cloud model time; the results are shown in Figure 11.

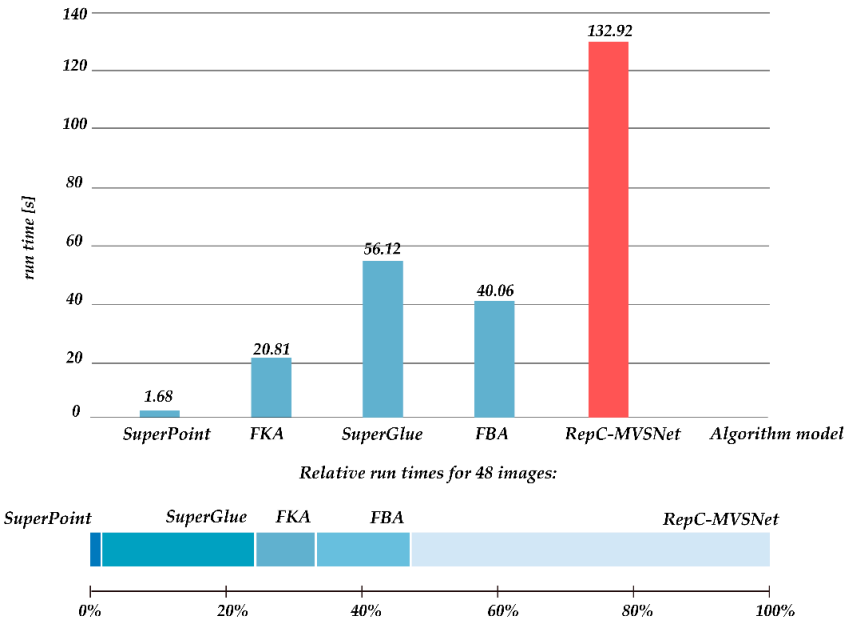

Figure 11. SfM-MVS system time consumption (at 48 images).

The time consumption of the whole 3D reconstruction process, the running time of SfM reconstruction, and the running time of MVS are not much different, in which the extraction, matching, and adjustment of key points account for about 47% of the total time consumption of the system, and the total time consumption of 3D reconstruction accounts for about 53% of the total time consumption. The MVS system runtime mainly depends on the design of the algorithm itself because the MVS system makes a quantitative constraint on the source images so that it does not grow in a factorial multiple with the number of images.

4.2.5. Depth Map for Wheat

The depth map is an intermediate bridge from stereo matching to point cloud generation. Each pixel stores the vertical depth value of the 3D object with respect to the viewpoint plane, and the 3D coordinates corresponding to each pixel point in the image in the spatial coordinate system can be obtained through the camera parameter matrix. The depth prediction results for our network are shown in Figure 12. From the individual depth maps, we can see that our reconstructed network works well for depth estimation at the global scale, can segment the wheat plant and background boundaries well using the difference in spatial location, can clearly present the overall extension of the wheat leaves, shows the 3D geometric spatial information, and has a high degree of hit for the real boundaries, but there is some fuzzy overlap phenomenon in the highly textured area, and there is leaf adhesion phenomenon.

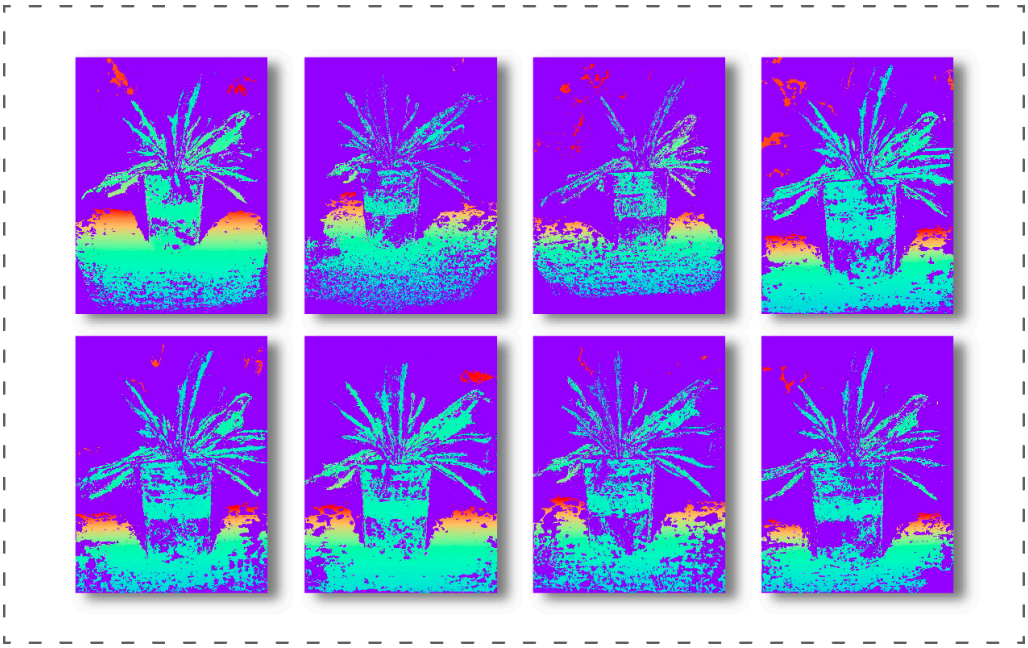

Figure 12. Wheat depth map.

4.2.6. Result of 3D Reconstruction for Wheat

The point cloud data based on multi-view reconstruction not only contains high-quality point cloud 3D information but also contains the color information of the original image data, which can better extract the phenotypic information of the overall reconstructed wheat and realize the spatial virtualization of the crop. From Figure 13, we can see that our dense point cloud model can clearly present the geometric structure and detailed texture of the whole wheat plant, and the color reproduction of the point cloud is basically consistent with the real scene, and there is no reconstruction mutilation phenomenon. At the same time, the overall point cloud is less noisy, and there is no misrepresentation of branches and leaves due to occlusion. The reconstructed wheat 3D point cloud data is about 200 MB, and the overall reconstruction data (including point cloud, depth map, camera pose parameters, etc.) is about 700 MB in size, which can be used as the dataset for subsequent 3D tasks.

At the same time, we compared the point cloud generation result of our model with that of the traditional method Colmap [38] (Figure 14). It can be seen that our model has a more complete and delicate processing of the wheat surface reconstruction information. The point cloud model reconstructed by Colmap has a phenomenon of emptiness and sparsity, which only reconstructs the original geometric features of wheat; the reconstruction time is 2 h. Our point cloud model does not have defects and vacancies, and the color saturation is higher, which truly reflects the three-dimensional details of wheat plants. In terms of reconstruction time, it only takes 5 min, reducing the overall reconstruction time to 4%.

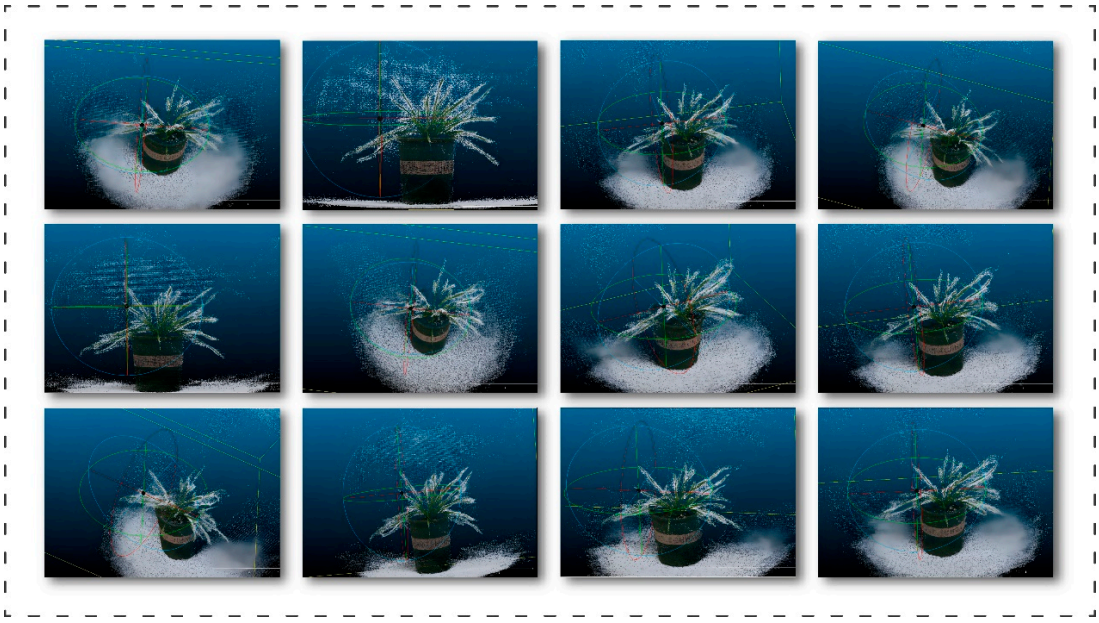

Figure 13. Multi-view 3D reconstruction results of wheat.

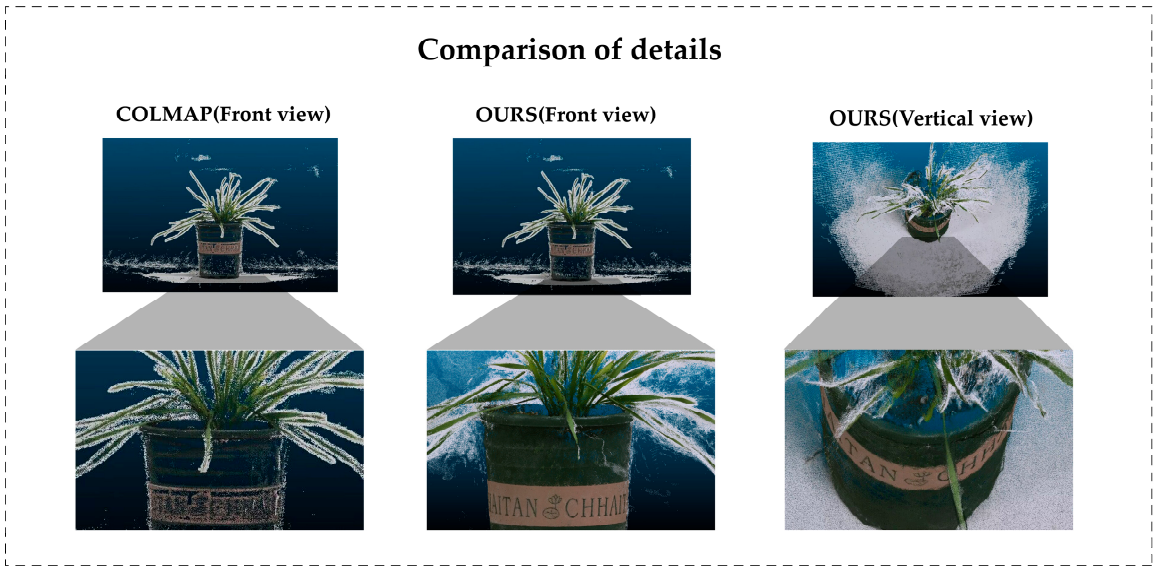

Figure 14. Comparison of wheat point cloud model reconstruction.

5. Discussion

In this study, we propose an integrated framework for the 3D reconstruction of wheat phenotypic features to solve the problems of wheat phenotypes that are difficult to analyze due to data masking and data distortion, the scarcity of existing wheat 3D data sets, the overly idealized existing 3D models, and the incomplete information of wheat when observing 2D images. It provides researchers with a complete 3D model of wheat so that they can observe the real wheat phenotype structure in more detail.

5.1. Contribution to Wheat 3D Point Cloud Data Generation

For wheat, its complex phenotype determines that it is difficult to obtain more wheat data information through 2D images, while point cloud data can better solve the data occlusion problem and then obtain more realistic analysis results through our 3D model of wheat growth. In the reconstruction task, we obtained a sparse point cloud containing information on key points of the wheat phenotype and a dense point cloud restoring

the structure of the wheat phenotype, and the 3D visualization model form of wheat morphology constructed in this way has a strong sense of realism. At the same time, the reconstructed point cloud can continue the task research of 3D target detection and 3D semantic segmentation, which provides a dataset basis with wide application prospects.

5.2. Contribution to Realizing Camera Pose Repositioning

The traditional camera calibration algorithm mainly uses the mathematical model based on the specific camera model and the corresponding relationship between the known space point and the image point and determines the camera intrinsics and camera extrinsics by minimizing the reprojection error, which is easy to have interfered by camera distortion and has a high probability of recalibration. To address this phenomenon, we used pixel-perfect SfM to train the depth features of images, optimize the key point matching results in different feature maps to obtain more accurate camera poses, solve the image alignment problem when the light and shooting position change, and correctly obtain the world coordinate system corresponding to the collected wheat images to achieve multi-view reconstruction effects.

5.3. Contribution to Self-Supervised 3D Model Construction for Wheat

Previously, 3D reconstruction of virtual crops was mostly based on structured light or LiDAR scanners with the former application scenarios being more ideal, while the latter mainly relied on expensive RGB-D cameras. The traditional 3D reconstruction algorithm is not only computationally intensive but also has a large gap between the modeling effect and the physical object; while the modeling using RGB-D cameras is highly accurate, the equipment is expensive and relies heavily on manual operation, which makes the equipment deployment difficult. While the core idea of the reconstruction algorithm using deep learning is to obtain the depth map of the scene, we use the RepC-MVSNet algorithm to obtain more 3D information by constructing a self-supervised method, combining depth rendering consistency, reference view synthesis consistency to reconstruct the loss function instead of the depth map, reconstructing to obtain a dense point cloud, and then 3D reconstruction. The method relies only on acquiring multi-view 2D images, which are extremely easy to deploy with conventional cameras. The method is not only cost-saving but also more scalable in scene applications compared to depth cameras.

5.4. Contribution to Agronomy Research

The method of multi-view point cloud generation and 3D reconstruction holds enormous potential in agricultural research and application. This method fills the gap of the multi-view, unsupervised reconstruction method in the field of agricultural research and also realizes the identification and evaluation of phenotypes of different varieties so as to screen and breed excellent varieties more effectively. On this basis, in genetic research and genotype–phenotype association, the phenotypic data obtained through three-dimensional reconstruction can be associated with genetic data to explore the relationship between genotype and phenotype. This is helpful to understand the effects of specific genes on the phenotypic traits of crops and provides the basis for molecular marker-assisted selection in the breeding process. In addition, the method of multi-view point cloud generation and three-dimensional reconstruction can realize the phenotypic evaluation of crops under adverse conditions, such as drought and high temperature. Through the three-dimensional reconstruction of the performance of different varieties under adverse conditions, we can evaluate their stress resistance and provide support for breeding varieties with strong adaptability. In future breeding research, researchers do not have to go to the breeding site frequently to observe, measure, and record the phenotypic data of wheat, and all the work can be done automatically by computer and give the results of phenotypic data analysis that can be referenced in a certain range. By providing more comprehensive and high-throughput phenotypic data of crops, this method is helpful to accelerate the breeding

process, cultivate better crop varieties and make positive contributions to agricultural production and food security.

6. Conclusions

This paper discusses a new idea of a phenotype analysis method for wheat crops in modern agricultural production, proposes a 3D reconstruction method for wheat based on SfM and MVS, and designs a point cloud generation network RepC-MVSNet. The method uses SfM to obtain camera parameter matrix information of wheat input image adaptively and makes the algorithm more accurate and robust by introducing various optimization and adjustment strategies combined with a self-supervision method for training. The algorithm is more accurate and robust and solves the problem of low reconstruction capability due to poor feature extraction in the previous 3D reconstruction process. Our proposed SfM reconstruction system of SuperPoint + SuperGlue + FKA + FBA improves the average track length to 7.1271 and reduces the minimum reprojection error to 1.1791, which successfully improves the matching accuracy of measurement results and reduces the overall wheat crop reconstruction error in complex scenes. At the same time, our model reconstruction time is 4% of the traditional Colmap reconstruction time, and the reconstructed point cloud details are highly restored, which provides a new idea for the 3D simulation prediction of plants. In future work, we can continue to improve the network model and enhance the feature extraction capability to realize an end-to-end 3D reconstruction system.

Author Contributions: Conceptualization, H.L.; methodology, H.L.; software, H.H. and Y.W.; validation, C.X., M.L. and H.H.; formal analysis, H.L.; investigation, H.L.; resources, H.L.; data curation, C.X.; writing—original draft preparation, H.L.; writing—review and editing, H.L., M.W. and J.L.; visualization, M.L.; supervision, M.W.; project administration, H.L.; funding acquisition, J.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Innovation and Entrepreneurship Training Program for College Students (Grant No. 202310626010), the Sichuan Province Department of Education (Grant NO. JG2021-464).

Data Availability Statement: The data in this study are available on request from the corresponding author.

Acknowledgments: We are grateful to Yongliang Ding for his instruction and guidance on the paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Lobos, G.A.; Camargo, A.V.; Del Pozo, A.; Araus, J.L.; Ortiz, R.; Doonan, J.H. Editorial: Plant Phenotyping and Phenomics for Plant Breeding. *Front. Plant Sci.* **2017**, *8*, 2181. [CrossRef] [PubMed]
2. Paproki, A.; Sirault, X.; Berry, S.; Furbank, R.; Fripp, J. A novel mesh processing based technique for 3D plant analysis. *BMC Plant Biol.* **2012**, *12*, 63. [CrossRef] [PubMed]
3. Wang, B.; Lin, C.; Xiong, S. Wheat Phenotype Extraction via Adaptive Supervoxel Segmentation. In Proceedings of the 2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Seoul, Republic of Korea, 16–19 December 2020; pp. 807–814.
4. Toda, Y.; Okura, F.; Ito, J.; Okada, S.; Kinoshita, T.; Tsuji, H.; Saisho, D. Training instance segmentation neural network with synthetic datasets for crop seed phenotyping. *Commun. Biol.* **2020**, *3*, 173. [CrossRef]
5. Lakshmi, S.; Sivakumar, R. *Plant Phenotyping Through Image Analysis Using Nature Inspired Optimization Techniques*; Intelligent Systems Reference Library; Springer: Cham, Switzerland, 2018.
6. Su, J.; Yi, D.; Su, B.; Mi, Z.; Liu, C.; Hu, X.; Xu, X.-m.; Guo, L.; Chen, W.H. Aerial Visual Perception in Smart Farming: Field Study of Wheat Yellow Rust Monitoring. *IEEE Trans. Ind. Inform.* **2021**, *17*, 2242–2249. [CrossRef]
7. Zhao, J.; Zhang, X.; Yan, J.; Qiu, X.; Yao, X.; Tian, Y.; Zhu, Y.; Cao, W. A Wheat Spike Detection Method in UAV Images Based on Improved YOLOv5. *Remote Sens.* **2021**, *13*, 3095. [CrossRef]
8. Mi, Z.; Zhang, X.; Su, J.; Han, D.; Su, B. Wheat Stripe Rust Grading by Deep Learning With Attention Mechanism and Images From Mobile Devices. *Front. Plant Sci.* **2020**, *11*, 558126. [CrossRef]
9. Gong, B.; Ergu, D.; Cai, Y.; Ma, B. Real-Time Detection for Wheat Head Applying Deep Neural Network. *Sensors* **2020**, *21*, 191. [CrossRef]

10. Hu, G.; Qian, L.; Liang, D.; Wang, M. Self-adversarial Training and Attention for Multi-task Wheat Phenotyping. *Appl. Eng. Agric.* **2019**, *35*, 1009–1014. [CrossRef]
11. Sandhu, K.S.; Lozada, D.N.; Zhang, Z.; Pumphrey, M.O.; Carter, A.H. Deep Learning for Predicting Complex Traits in Spring Wheat Breeding Program. *Front. Plant Sci.* **2021**, *11*, 613325. [CrossRef]
12. Kempthorne, D.M.; Turner, I.W.; Belward, J.A.; McCue, S.W.; Barry, M.D.; Young, J.; Dorr, G.J.; Hanan, J.; Zabkiewicz, J.A. Surface reconstruction of wheat leaf morphology from three-dimensional scanned data. *Funct. Plant Biol. FPB* **2015**, *42*, 444–451. [CrossRef]
13. Zhang, H.; Wang, Q.; Zhang, H.; Ji, Y.; Ma, X.; Xi, L. Wheat Three-Dimensional Reconstruction and Visualization System. *Appl. Mech. Mater.* **2012**, *195–196*, 1300–1307.
14. Chebrolu, N.; Läbe, T.; Stachniss, C. Spatio-Temporal Non-Rigid Registration of 3D Point Clouds of Plants. In Proceedings of the 2020 IEEE International Conference on Robotics and Automation (ICRA), Paris, France, 31 May–31 August 2020; pp. 3112–3118.
15. McElrone, A.J.; Choat, B.; Parkinson, D.Y.; MacDowell, A.A.; Brodersen, C.R. Using high resolution computed tomography to visualize the three dimensional structure and function of plant vasculature. *J. Vis. Exp. JoVE* **2013**, *74*, e50162.
16. Verboven, P.; Herremans, E.; Helfen, L.; Ho, Q.T.; Abera, M.K.; Baumbach, T.; Wevers, M.; Nicolaï, B.M. Synchrotron X-ray computed laminography of the three-dimensional anatomy of tomato leaves. *Plant J. Cell Mol. Biol.* **2015**, *81*, 169–182. [CrossRef]
17. Di Gennaro, S.F.; Matese, A. Evaluation of novel precision viticulture tool for canopy biomass estimation and missing plant detection based on 2.5D and 3D approaches using RGB images acquired by UAV platform. *Plant Methods* **2020**, *16*, 91. [CrossRef] [PubMed]
18. Fang, W.; Feng, H.; Yang, W.; Yang, W.; Duan, L.; Chen, G.; Xiong, L.; Liu, Q. High-throughput volumetric reconstruction for 3D wheat plant architecture studies. *J. Innov. Opt. Health Sci.* **2016**, *9*, 1650037. [CrossRef]
19. Qi, C.; Su, H.; Mo, K.; Guibas, L.J. PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2016; pp. 77–85.
20. Qi, C.; Yi, L.; Su, H.; Guibas, L.J. PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space. *arXiv* **2017**, arXiv:1706.02413.
21. Fan, H.; Su, H.; Guibas, L.J. A Point Set Generation Network for 3D Object Reconstruction from a Single Image. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2016; pp. 2463–2471.
22. Yang, Y.; Zhang, J.; Wu, K.; Zhang, X.; Sun, J.; Peng, S.; Li, J.; Wang, M. 3D Point Cloud on Semantic Information for Wheat Reconstruction. *Agriculture* **2021**, *11*, 450. [CrossRef]
23. Yao, Y.; Luo, Z.; Li, S.; Fang, T.; Quan, L. MVSNet: Depth Inference for Unstructured Multi-view Stereo. In Proceedings of the European Conference on Computer Vision, Munich, Germany, 8–14 September 2018.
24. Yao, Y.; Luo, Z.; Li, S.; Shen, T.; Fang, T.; Quan, L. Recurrent MVSNet for High-Resolution Multi-View Stereo Depth Inference. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 5520–5529.
25. Chen, R.; Han, S.; Xu, J.; Su, H. Point-Based Multi-View Stereo Network. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Republic of Korea, 27 October–2 November 2019; pp. 1538–1547.
26. Luo, K.; Guan, T.; Ju, L.; Huang, H.; Luo, Y. P-MVSNet: Learning Patch-Wise Matching Confidence Aggregation for Multi-View Stereo. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Republic of Korea, 27 October–2 November 2019; pp. 10451–10460.
27. Gu, X.; Fan, Z.; Zhu, S.; Dai, Z.; Tan, F.; Tan, P. Cascade Cost Volume for High-Resolution Multi-View Stereo and Stereo Matching. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 14–19 June 2020; pp. 2492–2501.
28. Yang, J.; Mao, W.; Álvarez, J.M.; Liu, M. Cost Volume Pyramid Based Depth Inference for Multi-View Stereo. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 14–19 June 2020; pp. 4876–4885.
29. Xu, H.; Zhou, Z.; Qiao, Y.; Kang, W.; Wu, Q. Self-supervised Multi-view Stereo via Effective Co-Segmentation and Data-Augmentation. *arXiv* **2021**, arXiv:2104.05374. [CrossRef]
30. Chang, D.; Bozic, A.; Zhang, T.; Yan, Q.; Chen, Y.; Süsstrunk, S.; Nießner, M. RC-MVSNet: Unsupervised Multi-View Stereo with Neural Rendering. In *European Conference on Computer Vision*; Springer Nature: Cham, Switzerland, 2022.
31. Huang, B.; Huang, C.; He, Y.; Liu, J.; Liu, X. M3VSNET: Unsupervised Multi-Metric Multi-View Stereo Network. In Proceedings of the 2021 IEEE International Conference on Image Processing (ICIP), Anchorage, AK, USA, 19–22 September 2021; pp. 3163–3167.
32. Lindenberger, P.; Sarlin, P.-E.; Larsson, V.; Pollefeys, M. Pixel-Perfect Structure-from-Motion with Featuremetric Refinement. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, BC, Canada, 11–17 October 2021; pp. 5967–5977.
33. Chum, O.; Matas, J.; Kittler, J. Locally Optimized RANSAC. In *Pattern Recognition. DAGM 2003. Lecture Notes in Computer Science*; Michaelis, B., Krell, G., Eds.; Springer: Berlin/Heidelberg, Germany, 2003; Volume 2781. [CrossRef]
34. LoweDavid, G. Distinctive Image Features from Scale-Invariant Keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110.

35. DeTone, D.; Malisiewicz, T.; Rabinovich, A. SuperPoint: Self-Supervised Interest Point Detection and Description. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Salt Lake City, UT, USA, 18–22 June 2018; Volume 726, pp. 337–33712.
36. Dusmanu, M.; Rocco, I.; Pajdla, T.; Pollefeys, M.; Sivic, J.; Torii, A.; Sattler, T. D2-Net: A Trainable CNN for Joint Detection and Description of Local Features. *arXiv* **2019**, arXiv:1905.03561.
37. Revaud, J.; Weinzaepfel, P.; Souza, C.R.d.; Pion, N.e.; Csurka, G.; Cabon, Y.; Humenberger, M. R2D2: Repeatable and Reliable Detector and Descriptor. *arXiv* **2019**, arXiv:1906.06195.
38. Schönberger, J.L.; Frahm, J.-M. Structure-from-Motion Revisited. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 4104–4113.