Computers, Materials & Continua
DOI: 10.32604/cmc.2022.027079
Article

Tech Science Press

Research on Multi-View Image Reconstruction Technology Based on Auto-Encoding Learning

Tao Zhang[1], Shaokui Gu[1], Jinxing Niu[1,*] and Yi Cao[2]

[1]School of Mechanical Engineering, North China University of Water Conservancy and Hydroelectric Power, Zhengzhou, 450045, China
[2]Department of Electrical and Computer Engineering, University of Windsor, Windsor, N9B 3P4, ON, Canada
*Corresponding Author: Jinxing Niu. Email: njx.mail@163.com
Received: 12 January 2022; Accepted: 02 March 2022

Abstract: Traditional three-dimensional (3D) image reconstruction method, which highly dependent on the environment and has poor reconstruction effect, is easy to lead to mismatch and poor real-time performance. The accuracy of feature extraction from multiple images affects the reliability and real-time performance of 3D reconstruction technology. To solve the problem, a multi-view image 3D reconstruction algorithm based on self-encoding convolutional neural network is proposed in this paper. The algorithm first extracts the feature information of multiple two-dimensional (2D) images based on scale and rotation invariance parameters of Scale-invariant feature transform (SIFT) operator. Secondly, self-encoding learning neural network is introduced into the feature refinement process to take full advantage of its feature extraction ability. Then, Fish-Net is used to replace the U-Net structure inside the self-encoding network to improve gradient propagation between U-Net structures, and Generative Adversarial Networks (GAN) loss function is used to replace mean square error (MSE) to better express image features, discarding useless features to obtain effective image features. Finally, an incremental structure from motion (SFM) algorithm is performed to calculate rotation matrix and translation vector of the camera, and the feature points are triangulated to obtain a sparse spatial point cloud, and meshlab software is used to display the results. Simulation experiments show that compared with the traditional method, the image feature extraction method proposed in this paper can significantly improve the rendering effect of 3D point cloud, with an accuracy rate of 92.5% and a reconstruction complete rate of 83.6%.

Keywords: Multi-view; image reconstruction; self-encoding; feature extraction

1 Preface

Emergence of 3D digital technology has greatly promoted the accurate modeling of 3D objects [1–3]. Traditional 3D digital technology will touch and disturb the reconstructed target in the process

of acquiring data, and acquisition time is relatively long. Development and application of non-contact three-dimensional object reconstruction technology has become an urgent problem in the field of accurate digitization of object surfaces.

Researchers usually use two 3D reconstruction methods. One of them is 3D point cloud of crop fruits directly obtained through sensors [4–6]. One method is to directly obtain 3D point cloud of object through the sensors [4]. Such methods can obtain high-precision object surface point clouds, but there are problems such as complicated and expensive equipment, large amount of data, low reconstruction efficiency, and large environmental lighting restrictions. Another method is to recover 3D phenotype point cloud of object surface through 2D images [6]. This type of method is simple to operate and low in price. The generated point cloud model often contains a lot of noise, and is greatly affected by the camera's calibration accuracy and ambient light in use. The reconstruction method of SFM, which acquires 2D image sequences and can automatically perform camera calibration, has great advantages. However, for the target object with complex structure, sparse point cloud obtained by SFM method contains less 3D information. Firstly, some operators are used to extract and match the features of the image. The purpose is to obtain high-quality effective features such as SIFT [7], speeded up robust features (SURF) [8], Oriented Fast and Rotated Brief (ORB) [9] features and so on. Then pose parameters of multi-eye camera and sparse point cloud of scene are obtained through motion recovery structure algorithm, and the scene is densely reconstructed through multi-eye stereo vision algorithm, and finally 3D model can be obtained by post-processing of dense point cloud.

At present, some scholars use deep learning for 3D reconstruction, but so far there is no network that can only input multi-eye images and directly output 3D point clouds or other 3D structure networks. This is because deep learning is more difficult to deal with spatial geometry problem, and whole process of 3D reconstruction uses a lot of spatial geometry theory. The current mainstream method is to use deep learning methods to replace certain steps in the whole framework of 3D reconstruction. For example, learned invariant feature transform (LIFT) [10] framework uses deep learning to replace feature extraction module, and some networks such as Bundle Adjustment networks (BA-Net) [11,12] use deep learning to estimate global parameters of camera and so on. The emergence of deep learning has made the accuracy of 3D reconstruction higher and higher. At present, the most widely used method of deep learning is to use deep learning for image depth estimation, including monocular, binocular and multi-eye depth estimation. The 3D reconstruction can be completed by mapping the estimated depth image to the 3D space to obtain a dense point cloud and performing information fusion. On the other hand, the use of deep learning can add semantic information to 3D reconstruction, such as semantic segmentation of point clouds and 3D semantic reconstruction of scenes. The image features of some objects are similar to background features, and it is difficult to detect them with traditional machine vision methods. The autoencoder can learn effective features of data from a large amount of unlabeled data, avoiding the problem that supervised learning network requires a large amount of high-quality labeled data for training. Chen et al. [13] proposed U-Net network, which fuses the feature map in encoding with the feature map of corresponding size in decoding through idea of shortcut connection, so as to recover more spatial information in the upsampling process. Liu et al. [14] designed a feature compression activation module through squeeze-and-excitation networks (SENet) network and constructed SE-Unet model to strengthen the network learning of image features. Ma et al. [15] proposed a new decoder structure, which introduces a self-attention mechanism to decode cascaded deep features and shallow features, reduces accuracy loss in the upsampling process. Cheng et al. [16] proposed a new semantic segmentation algorithm to adopt a dense layer structure in the network, use grouped convolution to speed up the calculation, and introduce an attention mechanism to improve segmentation effect. Current methods focus on

local features in 2D images and do not consider the connection and correlation of features between multi-view images.

To solve the above problems, multi-view 3D reconstruction algorithm based on self-encoding network is proposed, which extracts the features of 2D image through self-encoding network. Firstly, multiple images from different perspectives are collected, and extracted by convolutional neural network and its extended self-encoding network to get the shape features. Secondly, Fish-Net is used to replace U-Net to solve the end-to-end communication problem and reduce the problem of feature loss. Finally, GAN loss function is used to reduce the loss of image edges and achieve fine extraction of image features. Experiments show that the features extracted based on the method in the text have a higher degree of recognition than the artificial features.

2 Multi-View 3D Reconstruction Algorithm

Multi-view 3D reconstruction algorithm is shown in Fig. 1. The basic principle is: extract features of image and perform feature matching, calculate essential matrix according to matching points, and then perform incremental or global SFM algorithm to calculate the rotation matrix and translation vector of camera, and triangulate feature points to get sparse spatial point cloud.

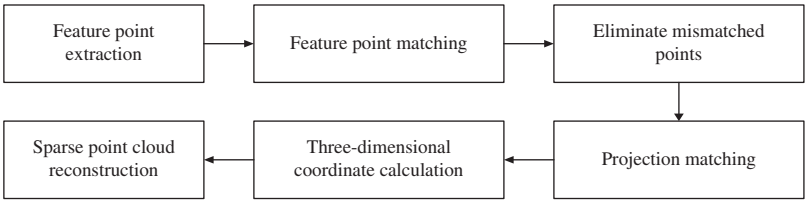

Figure 1: Multi-view 3D reconstruction algorithm

3 Image Feature Extraction Algorithm Based on Auto-Encoding Network

Feature extraction is one of the key technologies of multi-view image reconstruction, which purpose is to obtain high-quality effective features to prepare for image matching.

3.1 Auto-Encoding Algorithm

Autoencoder (AE) [17,18] is unsupervised artificial neural network model, which is widely used in data dimensionality reduction, noise reduction and sample reconstruction for data visualization, as shown in the Fig. 2. The autoencoder builds the U-Net network structure with the idea of sparse coding, generates low-dimensional features by encoding high-dimensional sample data, and uses low-dimensional and high-order coding features to reconstruct the original samples. The encoding network completes non-linear mapping of input data and outputs feature map. The feature map is then convolved and down-sampled to obtain multiple layers of hidden feature information. The decoding network uses feature map to reconstruct input data.

Since surface defects of objects are local anomalies in uniform textures, there are different characteristic representations between defects and background textures. Auto-encoding network is used to understand the representation of defect data and find the commonality of surface defects of object. Therefore, the problem of detecting defects on the surface of objects has become a problem of object segmentation. The encoder-decoder architecture is used to convert input defect image into pixel prediction mask.

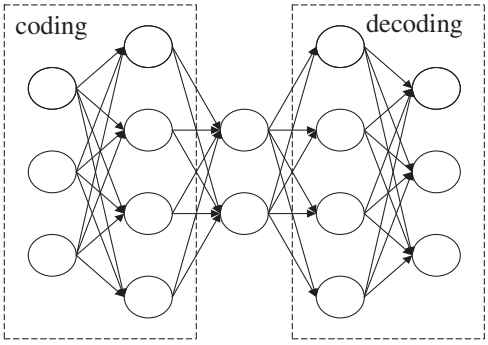

Figure 2: Structure of autoencoder

In the convolutional autoencoder network, mean-square error (MSE) [19] is often used as loss function. It is generally used to evaluate the pixel-level difference between two images, and to measure the pixel-level difference between reconstructed image and input image in the image reconstruction network. This function focuses on the global difference of the image and does not consider local texture features, therefore the inpainting model with MSE as the loss function performs better on regular texture samples than on the irregular texture.

3.2 Improvements to Auto-Encoding Deep Learning Networks

AE network is an end-to-end, simple and lightweight model, but the expression of model still does not achieve the desired effect. Therefore, Fish-Net is used instead of U-Net. In addition, only 2D feature point information does not provide best guidance. It is hoped that more information can be added to guide AE. AE uses a global loss function, but the weight of pixels inside object should be less than pixels near edges and surfaces.

3.2.1 Choice of Reconstruct Network Frames

In AE, two U-Nets in series are used to train a network of output image features end-to-end. U-Net uses the structure of "up/down sampling + jump connection", and the built neural network has the advantages of easy convergence and light weight. The deep network can easily obtain the gradient of the shallow network faster and retain the image pixel position information. However, the algorithm also has the problem that when multiple U-Nets work together on the same model, each U-Net directly cooperates poorly. Therefore, many other models based on U-Net have been improved, such as Fish NET and so on.

Fish-Net is an improvement on U-Net, as shown in the Fig. 3. Fish-Net consists of three parts: Tail, Body, Head. Tail uses the work of existing network structures to obtain deep low-resolution features from input image. Body contains upsampling and refinement extraction blocks, and obtains high-resolution features for high-level semantic information. Head contains down-sampling and refinement extraction blocks that preserve and refine the features obtained from these three parts. The refined features of last convolutional layer of Head are used for the final task decision. When multiple U-Nets are connected in series, there are skip connections between corresponding upsampling and downsampling within a single U-Net, but there are no skip connections between upsampling and downsampling in two adjacent U-Nets, so two paths between U-Nets may become a bottleneck for gradient propagation.

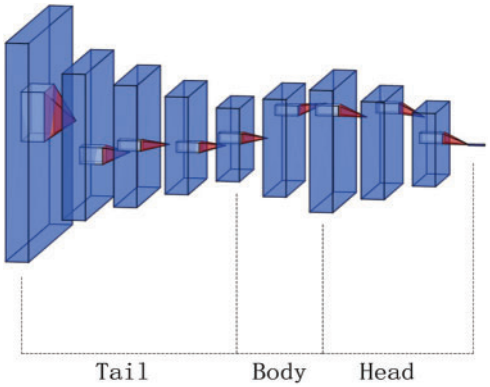

Figure 3: Structure of Fish-Net

Therefore, in addition to connecting the corresponding downsampling layer and upsampling layer in itself, Fish-Net also makes skip connections between each U-Net upsampling layer and the adjacent U-Net downsampling layer, so that the following U-Net is easy to feel the gradient of the previous U-Net. There are two convolutional blocks for upsampling and downsampling in Fish-Net, namely upsampling-reproducing block (UR-block) and downsampling-reproducing block (DR-block). For downsampling with stride 2, Fish-Net sets the convolution kernel size to 2×2, which resolves the overlap between pixels. To avoid weighted deconvolution with upsampling, Fish-Net chooses nearest neighbor interpolation for upsampling. Since the upsampling operation will dilute the input features at a lower resolution, Fish-Net also applies dilated convolutions, changing the data structure for retraining.

3.2.2 Improvement of Loss Function

Loss function is used to guide model training and play a key role in the training effect. GAN loss function used in this article is weighted by content loss and adversarial loss and expressed as:

$$L_{total} = L_{cont} + \lambda \cdot L_{adv} \tag{1}$$

where L_{total} is total loss; L_{cont} is content loss; L_{adv} is adversarial loss; λ is weight of adversarial loss function. Adversarial loss, which makes generated point cloud closer to actual point cloud by continuously optimizing generator and discriminator, is expressed as

$$L_{adv} = \frac{E}{S - p_{gen}(S)} [\log D(S)] + \frac{E}{B - p_{actual}(B)} [\log [1 - D(B)]] \tag{2}$$

where $p_{gen}(S)$ is distribution of generated point cloud; $p_{actuak}(B)$ is distribution of actual point cloud. The loss function makes generated point cloud more realistic in terms of visual effects. Content loss is expressed by a multi-stage loss function in this article. By considering of different edge features between blurred image and clear image in Stage1, L1 loss function is used in stage1. This function, which uses L1 gradient regularization [20–22] to constrain low-frequency feature detail information and retain more image edge information and structural details, is expressed as

$$L_{stage1} = \|B - S\|_1 + \beta \|\nabla B - \nabla S\|_1 \tag{3}$$

where $\|B - S_1\|$ is L1 loss; ∇ is gradient operator; β is weight of L1 gradient regularization. L2 loss function is used in Stage2 and Stage3, and helps solving problems such as lack of high-frequency

feature information during image generation. Compared with L1 loss, the image generated by L2 loss function training is more in line with the overall distribution of natural images. The function is expressed as

$$L_{stagei} = \frac{1}{2}\frac{1}{c_i w_i h_i}\|L_i - S_i\|^2 \quad i = 2, 3 \tag{4}$$

where L_i is output of generator model of i-th stage; S_i is clear image of i-th stage; c_i is number of channels in the i-th stage; w_i is width of i-th stage; h_i is height of i-th stage.

4 Experiment and Result Analysis

4.1 Experimental Environment

In order to evaluate effectiveness of the proposed algorithm, The extensive experiments are conducted on the published benchmark data set [23]. Experimental platform: high-speed visual processor (CPU i9-10900X, 3.7 GHz, 4.5 GHz Turbo, memory 64 GB DDR4, 32-bit Windows operating system). Algorithm uses matlab to write and debug the 3D reconstruction program, and adopts meshlab to display the final reconstruction results.

4.2 Network Training

Multi-view image feature extraction model structure includes: encoding network performs feature extraction and dimension compression on the input signal, decoding network is responsible for reconstructing the original signal, and classifier network uses the features extracted by encoding network to classify tasks.

The structure of each part of auto-encoding network is as follows.

(1) Input layer.

Input data is 500×600 two-dimensional data.

(2) Coding network.

Coding network consists of 3 convolutional layers alternating with 2 max pooling layers. The network performs feature extraction on the input data through convolution layer, and then uses maximum pooling layer to compress the features extracted by convolution layer to achieve feature dimension reduction.

(3) Decoding network.

Decoding network consists of 2 convolutional layers and 2 deconvolutional layers alternately. The network decodes the features extracted by coding network, and then uses deconvolution layer to map and expand the output feature map size to reconstruct the input signal.

(4) Activation function.

Activation function Rectified linear unit 6 (ReLU6) is used after convolutional layer and deconvolutional layer of the network. So ReLU6 solves the problem that corresponding weight cannot be updated because parts of input data fall into hard saturation region during training process.

(5) Global average pooling layer.

In order to reduce the feature dimension of coding network output, a "convolutional layer + global average pooling layer" structure follows output of coding network, and performs averagely

pooling to feature map output by coding network. Each feature map corresponds to a feature point, and finally these feature points are combined into a feature vector. Therefore, for input signals of different sizes, feature dimension extracted by network is fixed.

This paper conducts experiments on the proposed deep learning model based on the Pytorch framework. The optimizer adopts a periodic learning frequency of 10, an initial learning rate of 10, a learning rate decay rate of 0.000001, and a L2 regularization weight decay of 0.0001. Each batch randomly selects 200 image datas for training. Autoencoder model is first trained in two stages by the image dataset. The first stage reconstructs the original signal and saves the network parameters. In the second stage, a convolutional layer and a global average pooling layer follow the trained encoding network, and connect the classification network for training. The training process does not change the parameters of encoding network, but only updates the parameters of convolutional layer and classification network. Classifier is removed after training is complete. Therefore, output of the global average pooling layer is the feature extracted by auto-encoding network, and can be supplied to each classifier for classification. The article conducted comparative experiment on the size settings of epoch and batch-size, and finally determined that epochs = 10 and batch_size = 1024.

4.3 Multi-View Image Reconstruction Experiment

Take an angle of the image data set as an example, as shown in Fig. 4. Figs. 5 and 6 are used to represent two coordinate systems to extract image feature points.

Figure 4: Image data

Count the number of corresponding feature points of the image in the two images, as shown in Fig. 7.

Use the method in this article to generate a point cloud of a 3Dl object in a matlab environment, as shown in Fig. 8.

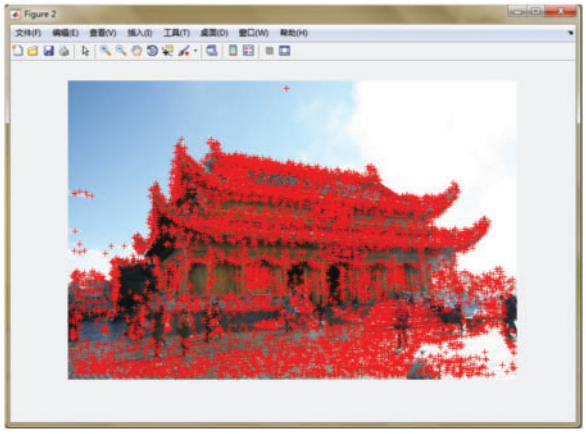

Figure 5: Reference image 1

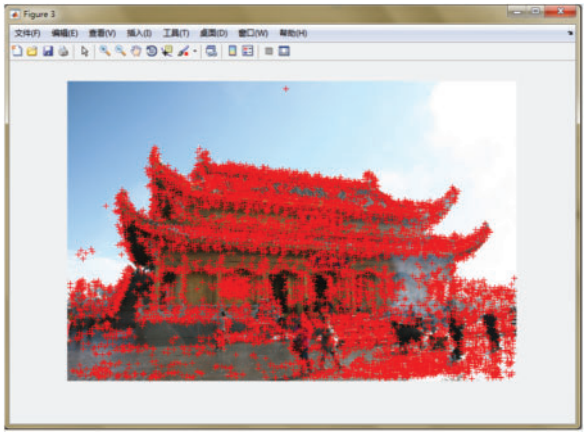

Figure 6: Reference image 2

Figure 7: Matching of different image feature points

The running time is about 15 s, and the result is better. It can be seen that, in the image, the more complex the color and shape features of the object, the less the pure color area, the better the reconstruction effect.

Problems encountered during matching operation:

(1) The network algorithm has a large amount of calculation, and the small CPU is easy to crash. When collecting images, it is necessary to control the image size or reduce the number of images on the premise of ensuring the quality.

(2) The interval angle between the two matching images should not be too small, which means that the camera does not move, and the matching cannot be processed. Therefore, it is necessary to ensure that there is a certain rotation change between the images.

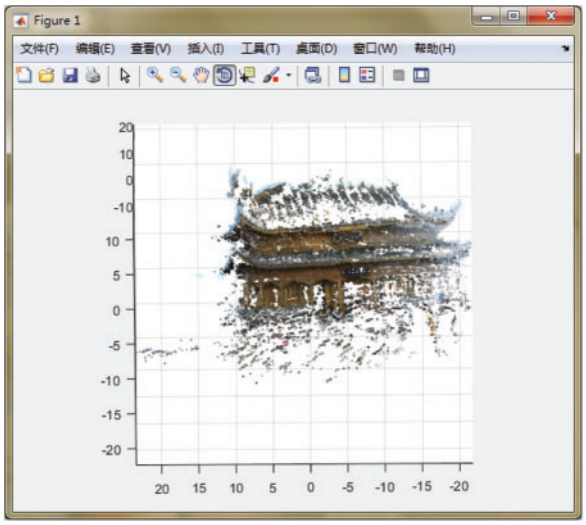

Figure 8: Generate point cloud diagram

The point cloud data generated in matlab is saved in dly format, imported into meshlab for rendering and display, and the final display result is shown in Fig. 9. The basic outline of the scene can be shown, but there are some problems in the simulation result. Because the background features are relatively few, the reconstruction effect of the background part is not good. The strong external light causes reflection on the surface of the object, which covers the color, and the matching effect is not good.

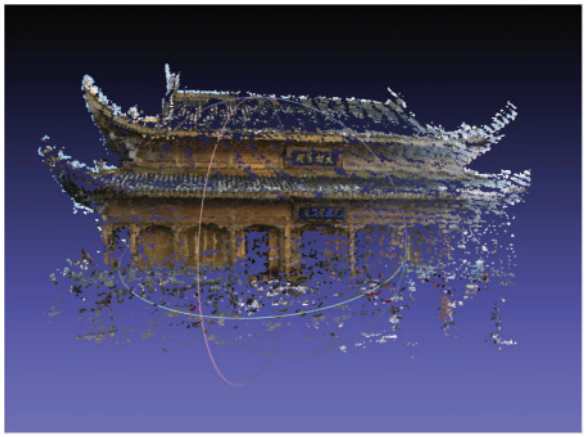

Figure 9: Meshlab displays the results

4.4 Algorithm Evaluation Index

In order to evaluate the effectiveness of the algorithm in this paper, which is compared with the effect of SIFT algorithm and Convolutional Neural Networks (CNN) network. Two indicators of

reconstruction accuracy and reconstruction completeness are used. The accuracy of reconstructed surface R measures closeness of reconstructed surface R to true surface G, and completeness of reconstructed surface R measures the extent to which true surface G is covered by reconstructed surface R. The comparison results are shown in Tab. 1. It can be seen that the algorithm in this paper is better than the other two algorithms. The traditional SIFT algorithm is highly sensitive to noise, and can easily lead to mismatches and convolutional neural networks. CNN adopts U-Net network structure, and directly splices the features of previous layer and the feature information obtained by corresponding downsampling during upsampling, resulting in the problem of feature loss. In this paper, after using the previous improved network, the parameters are greatly improved.

Table 1: Comparison of phase recovery of various methods

Algorithm	Precision	Completion
SIFT	71.2%	65.7%
CNN	84.2%	75.3%
Algorithm in this article	92.5%	83.6%

5 Conclusions

Aiming at the problem of inaccurate feature extraction and poor real-time performance in multi-view 3D reconstruction, an image feature extraction method based on auto-encoding network is proposed in this paper. The algorithm designs the auto-encoding network from two aspects of network structure and loss function. First, Fish-Net is used to replace U-Net, and skip connection of network layer is redesigned to improve the efficiency of network transmission. Secondly, GAN loss function is used to preserve more edge information and structural details of image. Experimental results show that reconstruction accuracy of the algorithm reaches 92.5%, and reconstruction integrity is 83.6%, which is better than the traditional U-Net network structure. However, in the environment of high reflective interference and weak features, the accuracy and real-time performance of this algorithm have not yet reached the highest level, so the model needs to be further optimized and improved in the next research:

(1) The network used in this paper is to initialize the parameters before training. It can improve the ability of network feature extraction in general by using a pre-trained model as an encoder.
(2) Fish-Net in this paper follows the network depth and width of U-Net in the design, however, it is not certain that this structure is the optimal solution. The next step is to explore the influence of network depth and width on the accuracy of target feature extraction.
(3) A more appropriate attention mechanism should be introduced to improve post-reconstruction processing optimization results.
(4) The simulation in this paper is only carried out on the marked public data set. The application range of the network can be further expanded for general extraction

Acknowledgement: The authors thank Dr. Jinxing Niu for his suggestions. The authors thank the anonymous reviewers and the editor for the instructive suggestions that significantly improved the quality of this paper.

Funding Statement: This work is funded by Key Scientific Research Projects of Colleges and Universities in Henan Province under Grant 22A460022, and Training Plan for Young Backbone Teachers in Colleges and Universities in Henan Province under Grant 2021GGJS077.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] Y. C. Wang, K. Liu and Q. Hao, "Robust active stereo vision using Kullback-Leibler divergence," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 3, pp. 548–563, 2012.
[2] M. Altalhi, S. U. Rehman, F. Alam, A. A. Alarood, A. U. Rehman *et al.,* "Computation of aortic geometry using mr and ct 3d images," *Intelligent Automation & Soft Computing*, vol. 31, no. 2, pp. 961–969, 2022.
[3] J. Zhang, X. Qi, S. H. Myint and Z. Wen, "Deep-learning-empowered 3d reconstruction for dehazed images in iot-enhanced smart cities," *Computers, Materials & Continua*, vol. 68, no. 2, pp. 2807–2824, 2021.
[4] M. Liu, M. Salzmann and X. M. He, "Discrete-continuous depth estimation from a single image," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, Columbus, OH, USA, pp. 716–723, 2014.
[5] X. R. Zhang, W. F. Zhang, W. Sun, X. M. Sun and S. K. Jha, "A robust 3-D medical watermarking based on wavelet transform for data protection," *Computer Systems Science & Engineering*, vol. 41, no. 3, pp. 1043–1056, 2022.
[6] C. Liu, J. M. Yang and D. Ceylan, "Plane net: Piece-wise planar reconstruction from a single RGB image," in *2018 IEEE Conf. on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, pp. 2579–2588, 2018.
[7] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.
[8] H. Bay, T. Tuytelaars and L. V. Gool, "SURF: Speeded up robust features," in *Proc. of the 9-th European Conf. on Computer Vision*, Graz, Austria, pp. 404–417, 2006.
[9] E. Rublee, V. Rabaud and K. Konolige, "ORB: An efficient alternative to SIFT or SURF," in *Int. Conf. on Computer Vision*, Barcelona, Spain, pp. 2564–2571, 2011.
[10] K. M. Yi, E. Trulls and V. Lepetit, "LIFT: Learned invariant feature transform," in *Proc. of the 16-th European Conf. on Computer Vision*, Amsterdam, Netherlands, pp. 467–483, 2016.
[11] C. L. Zheng, D. D. He and Q. G. Fei, "Improved digital image correlation method based on gray gradient regularization denoising," *Acta Optica Sinica*, vol. 38, no. 8, pp. 359–365, 2018.
[12] H. Zheng and D. Shi, "A multi-agent system for environmental monitoring using boolean networks and reinforcement learning," *Journal of Cyber Security*, vol. 2, no. 2, pp. 85–96, 2020.
[13] L. C. Chen, Y. Zhu and G. Papandreou, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proc. of the European Conf. on Computer Vision*, Munich, Germany, pp. 801–818, 2018.
[14] H. Liu, J. C. Luo, B. Huang, H. P. Yang, X. D. Hu *et al.,* "Building extraction based on SE-unet," *Journal of Geo-Information Science*, vol. 21, no. 11, pp. 1779–1789, 2019.
[15] H. Ma, H. J. Gao and T. Lei, "Semantic segmentation algorithm based on enhanced feature fusion decoder," *Computer Engineering*, vol. 46, no. 5, pp. 254–258+266, 2020.
[16] X. Y. Cheng, L. Z. Zhang and Q. Hu, "Fast semantic segmentation based on dense layerand attention mechanism," *Computer Engineering*, vol. 46, no. 5, pp. 247–252+259, 2020.
[17] F. N. Yuan, L. Zhang and J. T. Shi, "Summary of self-encoding neural network theory and application," *Journal of Computer*, vol. 42, no. 1, pp. 203–230, 2019.
[18] Y. J. Ren, Y. Leng, J. Qi, K. S. Pradip, J. Wang *et al.,* "Multiple cloud storage mechanism based on blockchain in smart homes," *Future Generation Computer Systems*, vol. 115, no. 2, pp. 304–313, 2021.

[19] S. Mei, H. Yang and Z. Yin, "An unsupervised-learning-based approach for automated defect inspection on textured surfaces," *IEEE Transactions on Instrumentation and Measurement*, vol. 67, no. 6, pp. 1266–1277, 2018.
[20] J. M. Fu, L. Chen and R. Zheng, "Survey of research on network attack detection based on GAN," *Netinfo Security*, vol. 19, no. 2, pp. 1–9, 2019.
[21] Y. J. Ren, F. J. Zhu, K. S. Pradip, T. Wang, J. Wang *et al.,* "Data query mechanism based on hash computing power of blockchain in internet of things," *Sensors*, vol. 20, no. 1, pp. 1–22, 2020.
[22] X. R. Zhang, X. Sun, X. M. Sun, W. Sun and S. K. Jha, "Robust reversible audio watermarking scheme for telemedicine and privacy protection," *Computers, Materials & Continua*, vol. 71, no. 2, pp. 3035–3050, 2022.
[23] Z. Y. Hu, "Three-dimensional reconstruction data set [EB]," 2010. http://vision.ia.ac.cn/data.