Applied Mathematics, Modeling and Computer Simulation
C. Chen (Ed.)

doi:10.3233/ATDE220085

3D Reconstruction of Digestive Tract Based on Wireless Capsule Endoscope Images

Nianfeng LI [a], Jia LU[c,1], Yuying WANG[b], Zhiguo XIAO[a] and Nan ZHAO[c]
[a] *College of Computer Science and Technology, Changchun University,China*
[b] *School of Cyber Security, Changchun University,China*
[c] *Graduate School of Changchun University, Changchun University,China*

Abstract. Based on the rapid development of 3D reconstruction technology and its expanding application in the direction of medical images, this paper proposed a research direction of 3D reconstruction of part of digestive tract based on single wireless capsule endoscope image by referring to Pixel2Mesh network. Experiments were carried out on the Data set of Kvasir-Capsule, and the results showed that compared with previous studies, the 3D model realized in this paper was more three-dimensional. At the same time, this paper provides a new idea for the development of 3D reconstruction technology in the direction of wireless capsule endoscope images.

Keywords. 3D reconstruction; wireless capsule endoscope; digestive tract; mesh model.

1. Introduction

The wireless capsule endoscope (hereinafter referred to as WCE) is the latest and non-invasive patient-friendly option for gastrointestinal examinations. There is no potential safety hazard of traditional intubation endoscopy (such as gastroscopy and colonoscopy), and the detection of the whole digestive tract can be realized [1]. The capsule endoscopy can be swallowed by the patient to enter the digestive tract and follow the peristaltic movement of the digestive tract to move. At the same time, the camera part inside the capsule endoscopy will take pictures of the digestive tract, and the captured images will be wirelessly transmitted to the receiving device outside the body through the sensor device inside the capsule. Doctors make diagnoses by observing the transmitted images.

Capsule endoscopic diagnosis also has some shortcomings: the field of vision taken by capsule lens is limited and not wide enough, and the viewing distance is short, it is difficult to detect far and larger lesions and the entire circumference of the expanded intestinal wall [1]. The structure of the inner wall of the digestive tract is complex, and blind areas such as folds or bends cannot be observed, resulting in a higher rate of missed detection [2] and so on. Among the methods to improve the information utilization of the capsule endoscopy, the use of the capsule endoscopy image for 3D reconstruction is

[1] Corresponding Author. Jia LU, Graduate School of Changchun University, Changchun University,China;E-mail:200701174@mails.ccu.edu.cn.

one of the most effective methods. The 3D reconstruction of the capsule endoscopy image is not only conducive to the doctor's focused observation and precise positioning of the lesion, but also Shorten the diagnosis time and lay the foundation for the next treatment.

There are many studies on WCE images 3D reconstruction of digestive tract, most of which use shadow algorithms [3], shadow technology [4], or a combination of image stitching and shadow technology [5], and contour segmentation combined with shadow technology [6] and other methods. All have achieved certain effects. However, due to the special environment of the digestive tract and the working principle of the capsule lens, it can only be used for single-view imaging. There is still a long way to go for the research on 3D reconstruction of part of digestive tract based on single-sheet WCE images.

In this paper, we refer to the deep learning architecture in [7] and propose some research directions for 3D reconstruction of part of digestive tract based on single-sheet WCE images. By this method, a single color WCE image can be used to generate a 3D shape of a 3D grid. Different from previous three-dimensional models based on capsule endoscopy images, convolutional neural network was used in this paper to extract perceptual features from the input images and generate the correct 3D shape of digestive tract from the image step by step, from coarse to fine. Compared with the previous technology, the image achieved by this method is more three-dimensional, which can provide a powerful reference for the doctor's diagnosis, and the application of capsule endoscopic images has broadened the application field of traditional 3D reconstruction technology, and continuous research and development in this field. It can also provide new ideas for the development of 3D reconstruction technology.

2. Related Work

3D reconstruction based on images is a common research in computer vision, medical image processing [8], and virtual reality [9]. In briefly, the following section introduces the research on the 3D reconstruction based on a single RGB image and the digestive tract reconstruction based on WCE images.

2.1. 3D Reconstruction of a Single Image

The three-dimensional reconstruction of object from a single image is an important direction in the field of computer vision. The traditional method of reconstructing 3D object from a single image requires prior knowledge and assumptions, and the reconstructed object is limited to a specific category, and it is difficult to complete a good reconstruction [10]. In recent years, great achievements have been made in 3d reconstruction of single image using deep learning technology. Fan H et al. [11] generated a direct form of output point cloud coordinates, and designed a conditional shape sampler, which can predict multiple credible 3D point clouds from a single input image. In addition, experiments have proved that this method is superior to other methods on a single image-based 3D reconstruction benchmark, and it also shows a powerful 3D shape completion capability. Wang N et al. [7] proposed an end-to-end deep learning architecture, Pixel2Mesh, which can generate the 3D shape of a triangular mesh from a single color image and better maintain the surface details of the object. Pixel2Mesh can represent a 3D grid in a graph-based convolutional neural network, using a network architecture similar to VGG-16 to extract perceptual features from the input image, using these perceptual features, gradually deform the ellipsoid and then generate the correct geometric shape. The strategy from coarse to fine is adopted to maintain the stability of the deformation process. Experiments have proved that this method can not only generate more detailed network models but also have higher accuracy of 3D shape estimation. Yu Z et al. [12] realized the 3D reconstruction based on a single image and single perspective. In the first stage, each pixel was mapped to the embedding space using CNN, where the embedding of pixels from the same plane instance was the same. Then, an effective mean shift clustering algorithm is used to group the embedding vectors into plane regions to obtain plane examples. In the second stage, the parameters of each plane instance are estimated by considering the pixel-level and instance-level consistency. And the effectiveness and efficiency of the method is verified on the public data set. In the field of medical image processing, the reconstruction of high-precision organ models is an effective method to improve the user's visual presence. Chen Q et al. [8] proposed a new 3D point cloud reconstruction model based on Morphing. Firstly, Mimics was used to obtain from a series of 2D CT images, then they use Morphing to perform 3D reconstruction of soft tissue through the sequential changes of the 3D surface model, and use nonlinear interpolation to fit the irregular shape of the model, which improves the accuracy of simulation. The soft tissues are more anastomosed.

However, 3D reconstruction based on deep learning also has many challenges: the accuracy of deep learning algorithms and models depends on a large amount of data sets, and due to the complexity of image processing, it often requires a lot of memory and computing time. At the same time, fine-grained 3D reconstruction of objects is also a difficult problem in 3D reconstruction based on deep learning.

2.2. Digestive Tract Reconstruction Based on WCE Images

Y.fan et al. [13] proposed an affine scale invariant feature Transform (SIFT) method for 3d reconstruction of intestinal wall surface, which was used to reconstruct the tubular surface of gastrointestinal tract in selected continuous frames. The relative motion of WCE is estimated by applying affine SIFT feature detector and descriptor to WCE image sequence. The Epipolar Geometry is used to further constrain the matched feature points to obtain an accurate 3D view. This provides more information and WCE image virtualization, which can effectively reduce the time it takes for doctors to detect abnormalities. Turan M et al. [5] first proposed a complete pipeline of 3D visualization of the stomach based on endoscopic images. The process is modular, including a preprocessing module, an image registration module, and a final 3D reconstruction module based on shadow shapes. The 3D mapping is mainly generated by a combination of image stitching and shadow shape technology, and iteratively updated frame by frame through the movement of the capsule in the stomach. And through experiments, it is proved that the RMS error is within the acceptable range.

3. Method and Results

3.1. Network Architecture

The overall network architecture diagram is shown in figure 1. The whole network is composed of image feature network and mesh deformation network. The image feature network is a 2D CNN, which extracts perceptual features from the input image, and uses the mesh deformation network to gradually deform the ellipsoid grid into the required 3D model. Our model learns to gradually deform and add details to the mesh model in a coarse-to-fine manner.

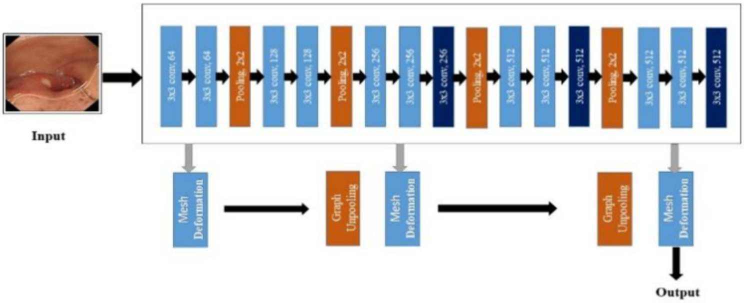

Figure 1. Network architecture: Through the mesh deformation structure part, the input single color image of the 3d shape is refined into a small mesh composed of 3D graphics.

Mesh Deformation: In order to generate a 3D mesh model consistent with the input image, the mesh deformation block pools the feature P extracted from the input image. The perception feature layer combines the input image features and the position of the vertex(C_{i-1}) of the given current model to obtain the aggregated perceptual features. The perceptual feature connection comes from the 3D shape feature on the vertices of the input graph(F_{i-1}), and the connection result is input into the graph-based ResNet (G-ResNet). G-ResNet generates new coordinates(C_i) and 3D shape features(F_i)of each vertex as the output of the mesh deformation module.

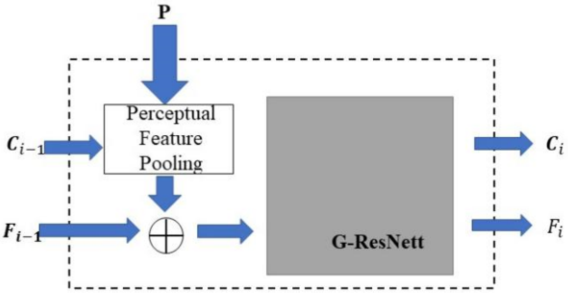

Figure 2. Mesh deformation block

The image feature network adopts a three-layer convolution VGG-16 architecture. The graph-based residual network (G-ResNet) is uniformly composed of 14 graph residual convolutional layers including 128 channels.

Graph unpooling layer: The unpooling layer is used to increase the number of fixed points in GCNN. This ensures that the network can start from a mesh with fewer vertices and gradually increase the vertices, which will produce better results while reducing memory costs. The non-pooling layer ensures the step-by-step refinement of the grid, making the generated 3D shape more suitable for real objects.

3.2. Losses

We defines four losses to constrain the output shape and deformation process. Chamfer loss (3-1) is used to constrain the position of the mesh vertices, normal loss (3-2) is used to enhance the consistency of surface normal, Laplace regularization (3-3) is used to maintain the deformation process the relative position between adjacent vertices, using edge length regularization (3-5) to prevent outliers.

In this section, use p to represent a vertex in the prediction grid, q to represent a vertex in the real grid, and $N_{(p)}$ to represent the adjacent elements of p. Laplace coordinates are as formula (3-4).

$$l_c = \sum_p min_q \parallel p - q \parallel_2^2 + \sum_q min_p \qquad (3-1)$$

$$l_n = \sum_p \sum_{q=\arg min_q(\parallel p-q \parallel_2^2)} \parallel < p - k, n_q > \parallel_2^2, s.t. k \in N_{(p)} \qquad (3-2)$$

$$l_{lap} = \sum_p \parallel \delta_p' - \delta_p \parallel_2^2 \qquad (3-3)$$

$$\delta_p = p - \sum_{k \in N_{(p)}} \frac{1}{\parallel N_{(p)} \parallel} k \qquad (3-4)$$

$$l_{loc} = \sum_p \sum_{k \in N_{(p)}} \parallel p - k \parallel_2^2 \qquad (3-5)$$

The total loss function is the weighted sum of the four loss functions (3-6). Where$\lambda_1 = 1.6e - 4, \lambda_2 = 0.3, \lambda_3 = 0.1$.

$$l_{all} = l_c + \lambda_1 l_n + \lambda_2 l_{lap} + \lambda_3 l_{loc} \qquad (3-6)$$

3.3. Results

We conduct experiments on the Kvasir-Capsule [14] dataset, which includes 4,741,507 images and 117 videos. We use 1500 images of this data set (1000 images in the training set and 5000 images in the test). The experimental results are shown in figure 3. It can be seen from the figure that the method used in this paper is a good way to transform the structure of the digestive tract into a 3D shape.

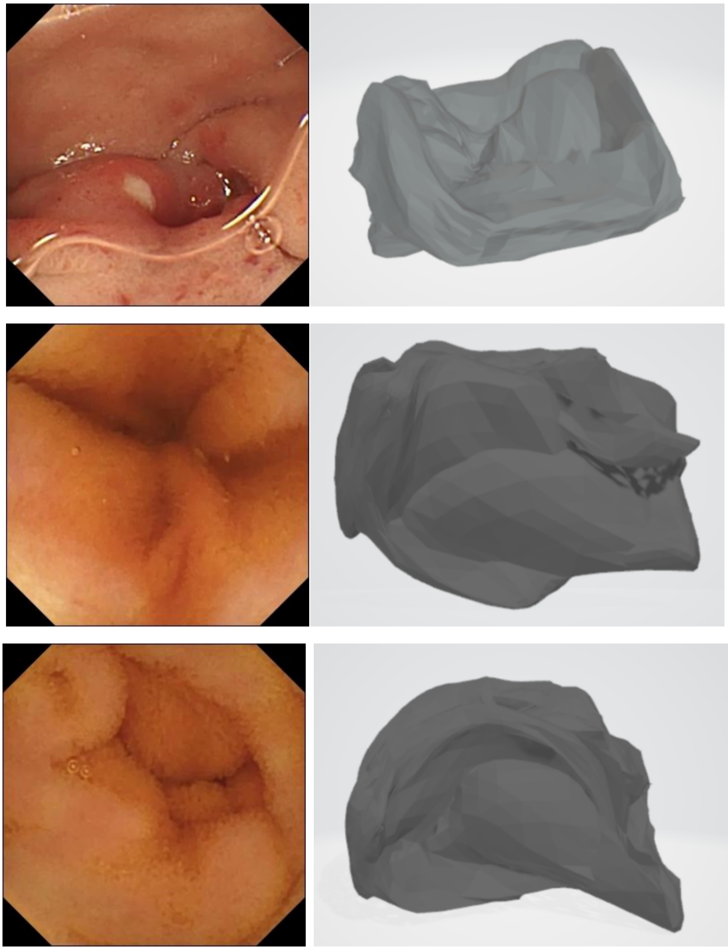

Figure 3. Experimental results

4. Conclusion

In this paper, referring to the network architecture in [7], the method of extracting three-dimensional triangular mesh from a single image was used to successfully transform the digestive tract structure contained in a single WCE images into the corresponding three-dimensional shape. The method of refining the surface mesh of the 3D model from coarse to fine makes the model more suitable for the real shape. We believe that the 3D mesh model based on the capsule endoscopy image will be an important direction for the 3D reconstruction of the digestive tract based on the capsule endoscopy. We hope that the contribution of this work can support subsequent research and further advance the development of this field.

Acknowledgements

The work was supported by the project plan of science and technology development center of the Ministry of Education (No. 2020ITA07031) and in part by the Jilin Science and Technology Development Plan Project (Project No. 20200404221YY).

References

[1] Xiao Z, Feng L N. A Study on Wireless Capsule Endoscopy for Small Intestinal Lesions Detection Based on Deep Learning Target Detection[J]. IEEE Access, 2020, 8: 159017-159026.

[2] Yanling Cao. Wireless Capsule endoscope image 3d reconstruction based on spline interpolation [D]. Harbin Institute of Technology,2015

[3] Koulaouzidis A, Karargyris A, Giannakou A, et al. The Use of Three-Dimensional Reconstruction Software in Oesophageal Capsule Endoscopy: A Pilot Study from Edinburgh[J]. Global Journal of Gastroenterology & Hepatology, 2014, 2(3): 84-91.

[4] Zhao Q, Meng M Q H. 3D reconstruction of gi tract texture surface using capsule endoscopy images[C]//2012 IEEE International Conference on Automation and Logistics. IEEE, 2012: 277-282.

[5] Turan M, Pilavci Y Y, Jamiruddin R, et al. A fully dense and globally consistent 3d map reconstruction approach for gi tract to enhance therapeutic relevance of the endoscopic capsule robot[J]. arXiv preprint arXiv:1705.06524, 2017.

[6] Prasath V B S, Figueiredo I N, Figueiredo P N, et al. Mucosal region detection and 3D reconstruction in wireless capsule endoscopy videos using active contours[C]//2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE, 2012: 4014-4017.

[7] Wang, N. , Zhang, Y. , Li, Z. , Fu, Y. , Liu, W. , & Jiang, Y. . (2018). Pixel2Mesh: Generating 3D Mesh Models from Single RsGB Images.. abs/1804.01654, 55-71.

[8] Cheng Q, Sun P, Yang C, et al. A morphing-Based 3D point cloud reconstruction framework for medical image processing[J]. Computer methods and programs in biomedicine, 2020, 193: 105495.

[9] Sra M, Garrido-Jurado S, Schmandt C, et al. Procedurally generated virtual reality from 3D reconstructed physical space[C]//Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology. 2016: 191-200.

[10] Fu K, Peng J, He Q, et al. Single image 3D object reconstruction based on deep learning: A review[J]. Multimedia Tools and Applications, 2021, 80(1): 463-498.

[11] Fan H, Su H, Guibas L J. A point set generation network for 3d object reconstruction from a single image[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 605-613.

[12] Yu Z, Zheng J, Lian D, et al. Single-image piece-wise planar 3d reconstruction via associative embedding[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 1029-1037.

[13] Fan Y, Meng M Q H. 3D reconstruction of the WCE images by affine SIFT method[C]//2011 9th World Congress on Intelligent Control and Automation. IEEE, 2011: 943-947.

[14] Smedsrud P H, Thambawita V, Hicks S A, et al. Kvasir-Capsule, a video capsule endoscopy dataset[J]. Scientific Data, 2021, 8(1): 1-10.