An effective approach for CT lung segmentation using region growing

Xi Yang1*, Guanyu Xu1 and Teng Zhou2

1 School of Information Science and Technology, Guangdong University of Foreign Studies, Guangzhou, Guangdong, 510006, China
2 College of Engineering, Shantou University, Shantou, Guangdong, China
* Corresponding author’s e-mail: xi_yang@gdufs.edu.com

Abstract. X-ray is an important means of detecting lung diseases. With the increasing incidence of lung diseases, computer-aided diagnosis technology is of great significance in clinical treatment. It has become a hot research direction to use computer-aided diagnosis to recognize chest radiography images, which can alleviate the uneven status of regional medical level. For clinical diagnosis, medical image segmentation can enable users to timely obtain the target region they are interested in and analyze it, which is significant to be used as an important basis for auxiliary research and judgment. In this case, a region growing algorithm based on threshold presegmentation is selected for lung segmentation, which integrates image enhancement, threshold segmentation, seed point selection and morphological post-processing, etc., to improve the segmentation effect, which also has certain reference value for other medical image processing.

1. Introduction

China is an area with a high incidence of lung cancer. In recent years, the incidence of liver cancer has been on the rise, and the mortality rate has also increased, and it has entered the forefront of cancer mortality rankings. [1] With the increase in the number of lung cancer patients, the number of lobectomy operations has also increased, which also puts more pressure on doctors to perform lung image analysis. Lung region segmentation is of great significance to image analysis. The traditional approach is to manually circle the slices of each axis in the lung image by doctors to achieve this. This increases labor costs and brings a certain degree of subjectiveness to the accuracy of segmentation. For this reason, the automatic segmentation technology of lung images has practical application value and is of great help to the improvement of the overall image analysis efficiency[2].

Medical CT image processing is mainly to study the relationship between organs and tissues in medical images and conduct pathological analysis. Therefore, the precise segmentation and positioning of the area that doctors pay attention to in CT images with the help of computer and image processing technology is a key step in medical image processing, which is of great significance to assisting doctors in pathological research and judgment in clinical diagnosis[3-5].

Medical imaging has strong timeliness and science, and is an important reference for clinical diagnosis. With the continuous development of image segmentation technology, a large number of segmentation algorithms have emerged, many of which have been applied and developed in combination with the actual needs of medical diagnosis. In the actual application process, image segmentation, as one of the most commonly used modules in diagnostic analysis, plays an increasingly
important role. Region growing segmentation is a classic image segmentation algorithm. Based on the idea of serial regions, it extracts connected regions with the same characteristics to obtain a complete target edge to achieve the segmentation effect. In this case, medical image segmentation is performed by the region growing method and combined with different self-processing methods to improve the effect, and an effective lung image segmentation method is obtained.

2. Related works

The commonly used digital image processing methods in the literature include thresholding, wavelet, region growing, active contour and watershed[6]. However, image processing still has many challenges in the medical field, especially because the size of the lungs constantly changes during clinical examinations. In various studies, methods for pathological detection and prediction have also been widely explored. Researchers from Norris Cotton Cancer Center developed and trained a new machine learning model[7]. The results show that the image results identified by the computer model are basically consistent with the subjective assessment results made by the pathologist. In order to build a reliable deep learning model trained and tested on a large-scale data set, the model [7] constructed a public COVID-19 CT data set, which contains 1186 lung cancer CT images. In addition, synthetic or real chest CT images are used to classify various deep learning models for COVID-19 and non-COVID-19. Through comparison, all models have achieved good results in the accuracy, recall, recall and F1 scores of synthetic and real COVID-19 CT images, which proves the reliability of the synthetic data set. Public data sets and deep learning models help to develop accurate and effective COVID-19 diagnostic tests. [8] The study conducted a comparative analysis of various segmentation methods of lung cancer CT images. The model proposed by the study [9] used the modified maximum transverse diameter method to mark a hypothetical lung nodule. This improvement includes the use of a set of appropriately sized overlapping spheres to approximate the shape of the nodule. The algorithm embedded in the model also groups the markings of the same lesion by different readers. The developed cluster model simplifies the collaboration and crowdsourcing creation of image libraries and improves time efficiency. Our proof-of-concept data set provides a valuable source of medical imaging data for training CAD algorithms aimed at early detection of lung nodules.

3. Proposed method

3.1. Threshold segmentation

Threshold segmentation is the most classic segmentation technique, and it is also the simplest and most practical. In many cases, there are differences in gray values between the target area and the background area or different areas in the image. The core of the threshold segmentation method is how to find the appropriate threshold. The most commonly used threshold method is based on the gray histogram, such as the maximum between-class variance method (OTSU)[10]. The selection of the threshold is generally based on the local statistical information of the image, such as local variance, local contrast, and surface fitting threshold. However, medical images generally contain multiple different types of regions, and how to select a suitable threshold value for segmentation from them is still a big problem for medical image threshold segmentation. As shown in Figure 1, the lung image segmentation result obtained by directly applying threshold segmentation contains more noise and over-segmentation.
3.2. Regional growing

Region Growing is essentially a process of combining seed pixels or sub-regions through predefined similarity calculation rules to obtain a larger region. First, select the seed pixel or sub-region as the target position; then, merge the adjacent pixels or regions that meet the similarity conditions to the target position, and loop to achieve the gradual growing of the region; finally, if there is no point or small region that can continue to be merged, stop and output. Among them, the similarity calculation rules can include gray value, texture, color and other information.

In the absence of prior knowledge, the region growing method finds the possibility of optimal segmentation through rule merging strategy. It has the characteristics of simplicity and efficiency, and has a better segmentation effect for more uniformly connected targets. However, the region growing method generally requires manual selection of seed points or sub-regions, which tends to lack objectivity; moreover, the region growing method is more sensitive to noise, which may cause problems such as holes and noise in the segmentation results. In addition, it is a serial algorithm. When the target is large, the segmentation speed is slower. Therefore, when designing the algorithm, it is necessary to improve the efficiency as much as possible.

As shown in Fig.2, the direct application of the region growing method requires manual selection of seed points, and there are many holes and noise edges in the output liver segmentation results, which also brings certain interference to the subsequent diagnostic analysis. For this reason, consider Combined with threshold segmentation, seed points are automatically selected and morphological post-processing is applied to remove holes and noise.

3.3. Region growing based on threshold pre-segmentation

The direct application of threshold segmentation algorithm to liver image is prone to over-segmentation problem, which is to segment a large number of other areas connected with the liver. If the region growing algorithm is directly applied, the seed points need to be manually selected, and the segmentation results are likely to contain problems such as holes and noise. Therefore, the approximate area of the liver is pre-positioned by threshold segmentation, and the seed point is selected according to the default position of the liver, and the binary image obtained after the region growing segmentation is subjected to morphological post-processing, and finally a complete liver target is obtained to achieve segmentation.
4. Results and discussion

4.1. Data set and model training
The data set is from the CT image data set used in the 2017 kaggle competition to segment lung images from CT images and identify lung volume cases. The samples included 267 cases. All data is provided by K Scott Mader and is open to use. The images in the data set have been artificially segmented lungs and measured in 2D/3D. This data set is the second version (2017.4.26) of the case upload and update. We only uses 2D images.

Take 50% of the initial data set as the test set. Using the holdout method, we divide the initial data set into two parts: a training set and a test set. The training set is used for model training, and the test set is used for performance evaluation.

4.2. Validity analysis of segmentation model
As shown in Figure 4, we get the gray histogram of each picture and the automatically selected threshold. According to the manual selection principle described above, the bottom of the double peak can be roughly 100. It is obvious that the segmentation effect of the threshold value of 100 is better than that of the threshold value of 150.
The area growing method mainly considers the relationship between pixels and their spatial neighborhood pixels. At the beginning, determine one or more pixels as seeds, and then grow the area according to the similarity criterion to gradually generate a spatial area with a certain uniformity, merge adjacent pixels or areas with similar properties to gradually increase the area until there is no For points or other small areas that can be merged, the similarity measure includes information such as average gray value, texture, and color.

| Evaluation Metrics | Table 1. Respective metric results |
|--------------------|----------------------------------|
|                    | MAE   | MSE   | RMSE  | ACC   |
| Coarse tree        | 0.868 | 0.858 | 0.801 | 0.853 |
| Matern 5/2 GRP     | 0.962 | 0.952 | 0.939 | 0.918 |
| Linear SVM         | 0.917 | 0.922 | 0.823 | 0.913 |
| Boosted trees      | 0.873 | 0.938 | 0.819 | 0.769 |

It can be seen from Table 1 that this segmentation model can achieve higher accuracy with Matern 5/2 GRP and linear SVM classification, but Matern 5/2 GRP is higher than SVM in MAE, MSE and RMSE, so the best model is based on threshold Pre-segmented area growing combined with Matern 5/2 GRP for lung cancer diagnosis.

This case is aimed at the application of lung image segmentation. It uses threshold pre-segmentation for automatic seed point selection, region growing for image segmentation, and morphological post-processing to improve the segmentation effect, while reducing iteration steps and time complexity. Therefore, the idea of combining the threshold method and the region growing method, combining the advantages of the two is feasible.
5. Conclusion
In this paper, we can find that the combination of the threshold method and the region growing method improve the segmentation effect. The method assists the pathologist in the image inspection work, reduce the doctor's diagnosis pressure, and improve the accuracy of the automatic segmentation of CT images.

Acknowledgments
This work supported by the Guangzhou Scientific and Technological Plan Project (No.202102021098), NSFC (No. 61902232), the Natural Science Foundation of Guangdong Province (No. 2018A030313291, 2019A1515010943).

References
[1] A.Gusnanto, P.Tcherbeniakov, F.Shuweihdi, M.Samman, et al. (2015) Stratifying tumour subtypes based on copy number alteration profiles using next generation sequence data, Bioinformatics, Vol.31, Issue 16, pp.2713-2720
[2] P.Li, X.Jiang, C.Kambhamettu, et al. (2018) Compound image segmentation of published biomedical figures, Bioinformatics, vol.34, Issue 7, pp.1192-1199
[3] V.I.Jurtz, A.R.johansen, M.Nielsen, et al. (2017) An introduction to deep learning on biological sequence data: examples and solutions, Bioinformatics, vol.33, Issue 22, pp.3685-3690
[4] D.Osumi-Sutherland, S.Reeve, C.J.Mungall, et al. (2012) A strategy for building neuroanatomy ontologies, Bioinformatics, vol.28, Issue 9, pp.1262-1269
[5] A.D.Anjos, A.L.B.Moller, B.K.Erbsoll, et al. (2011) New approach for segmentation and quantification of two-dimensional gel electrophoresis images, Bioinformatics, vol.27, Issue 3, pp.368-375
[6] Y.Lin, C.Lin, Y.Tsai, et al. (2010) A spectral graph theoretic approach to quantification and calibration of collective morphological differences in call images, Bioinformatics, vol.26, Issue 12, pp.i29-i37
[7] H.Jiang, S.Tang, W.Liu, et al. (2021) Deep learning for COVID-19 chest CT (computed tomography) image analysis: A lesson from lung cancer, Computational and Structural Biotechnology Journal, vol.19, pp.1391-1399
[8] R.Indumathi, R.Vasuki, (2021) Segmentation of Lung Cancer from CT Image- A Comparative Analysis, Materialstoday Proceedings, in press
[9] S.P.Morozov, V.A.Gombolevskiy, A.B. Elizarov, et al. (2021) A simplified cluster model and a tool adapted for collaborative labeling of lung cancer CT scans, Computer Methods and Programs in Biomedicine, vol.206, 106111
[10] G.E.Hoffman, J.Bendl, K.Girdhar, et al. (2020) Decorate: differential epigenetic correlation test, Bioinformatics, vol.36, Issue 9, pp.2856-2861
[11] O.Demirkaya, M.H.Asyali, M.M.Shoukri, Segmentation of cDNA Microarray Spots Using Markov Random Field Modeling, Bioinformatics, vol.21, Issue 13, pp.2994-3000
[12] S.M.Teo, Y.Pawitan, V.Kumar, et al. (2011) Multi-platform segmentation for joint detection of copy number variants, Bioinformatics, vol.27, Issue 11, pp.1555-1561
[13] S.Basu, D.Pewczynski, S.Saha, et al. (2016) 2dSpAnsemiautomated 2-d segmentation, classification and analysis of hippocampal dendritic spine plasticity, Bioinformatics, vol.32, Issue 16, pp.2490-2498