Computerized Medical Imaging and Graphics 63 (2018) 1–8

Contents lists available at ScienceDirect

Computerized Medical Imaging and Graphics

journal homepage: www.elsevier.com/locate/compmedimag

Multiple supervised residual network for osteosarcoma segmentation in CT images

Rui Zhang[a,1], Lin Huang[a,b,1], Wei Xia[a], Bo Zhang[c], Bensheng Qiu[b], Xin Gao[a,*]

[a] Suzhou Institute of Biomedical Engineering and Technology, Chinese Academy of Sciences, Suzhou, China
[b] University of Science and Technology of China, Hefei, China
[c] The Second Affiliated Hospital of Soochow University, Suzhou, China

ARTICLE INFO

Keywords:
Osteosarcoma segmentation
Deep residual network
Multiple supervised networks

ABSTRACT

Automatic and accurate segmentation of osteosarcoma region in CT images can help doctor make a reasonable treatment plan, thus improving cure rate. In this paper, a multiple supervised residual network (MSRN) was proposed for osteosarcoma image segmentation. Three supervised side output modules were added to the residual network. The shallow side output module could extract image shape features, such as edge features and texture features. The deep side output module could extract semantic features. The side output module could compute the loss value between output probability map and ground truth and back-propagate the loss information. Then, the parameters of residual network could be modified by gradient descent method. This could guide the multi-scale feature learning of the network. The final segmentation results were obtained by fusing the results output by the three side output modules. A total of 1900 CT images from 15 osteosarcoma patients were used to train the network and a total of 405 CT images from another 8 osteosarcoma patients were used to test the network. Results indicated that MSRN enabled a dice similarity coefficient (DSC) of 89.22%, a sensitivity of 88.74% and a F1-measure of 0.9305, which were larger than those obtained by fully convolutional network (FCN) and U-net. Thus, MSRN for osteosarcoma segmentation could give more accurate results than FCN and U-Net.

1. Introduction

Osteosarcoma is a type of extremely dangerous primary malignant bone tumor with poor prognosis. Its incidence accounts for about 0.2% of that of human malignant solid tumors (Meyers et al., 2008; Moore and Luu, 2014; Ottaviani and Jaffe, 2009). Accurate segmentation of osteosarcoma region from CT image is very important for preoperative planning of neoadjuvant chemoradiation therapy and postoperative assessment of therapeutic efficacy (Bacci et al., 2006; Ritter and Bielack, 2010; Luetke et al., 2014; Kim et al., 2008). However, manually outlining tumor region is a time-consuming and arduous task. In addition, the outlining results can be affected by radiologists' subjective experiences, the environment, etc., indicating poor repeatability of the results (Mandava et al., 2010). Thus, the automatic segmentation of osteosarcoma region is urgently needed in clinic. In fact, there are two factors affecting the accuracy of automatic tumor segmentation. First, the location, structure, size and shape of tumors vary with patients. Second, tumors have high heterogeneity (Arndt and Crist, 1999). The distribution of density in tumor area in CT image is not even and the density between tumor tissue and surrounding normal tissue can hardly be differentiated.

Unsupervised clustering-based method is most frequently used for osteosarcoma segmentation in CT images. Mandava et al. (2010) proposed a dynamic clustering algorithm based on the harmony search hybridized with fuzzy c-means to automatically segment osteosarcoma region in MRI images. In their work, a novel harmony search operator was introduced to support the selection of empty decision variables in the harmony memory vector. Fuzzy clustering method (FCM) was incorporated in the method to refine segmentation results. Chen et al. (2012) developed a segmentation method based on hybrid relative fuzzy connectedness. Confidence connected method was first used to produce a rough segmentation of osteosarcoma in CT images. Then the mean and variance values obtained from the rough regions were used as the initial values for fuzzy connectedness algorithm. Finally, fuzzy connectedness method was used to refine the rough segmentation results. Li et al. put forward a fuzzy connectedness method for the segmentation of osteosarcoma in multimodal MRI images (Ma et al., 2005). The Unsupervised clustering-based method has high computational

* Corresponding author at: Suzhou Institute of Biomedical Engineering and Technology, Chinese Academy of Sciences, Suzhou, Jiangsu, 215163, China.
E-mail addresses: xingaosam@yahoo.com, gaox@sibet.ac.cn (X. Gao).
[1] These authors are primary authors and contributed equally.

https://doi.org/10.1016/j.compmedimag.2018.01.006
Received 2 August 2017; Received in revised form 6 December 2017; Accepted 9 January 2018
0895-6111/

efficiency and is easy to implement. However, it is very susceptible to noise in the image and algorithm initialization. Thus, it can only be used for cases in which the tumor to normal contrast is large as well as for images with regular lesion pattern. In other words, this clustering method is not suitable for the segmentation of osteosarcoma in complicated CT images.

Supervised machine learning method is another method for the segmentation of osteosarcoma in CT images. In this method, image segmentation is considered as the classification of pixels (or voxels) in image. First, a certain number of hand-craft features such as texture features, wavelet features, morphologic features, etc. are extracted from the image. Then these features are used to train the classifier parameter. Finally, the classifier for segmentation is obtained. Chen et al. (2013) proposed an approach based on Zernike moment and support vector machine (SVM) for the segmentation of osteosarcoma in MRI images. Frangi et al. (2001) applied neural network method to segment osteosarcoma in MRI perfusion images. They first built a pharmacokinetic model for the MRI perfusion images and then extracted pharmacokinetic features of MRI perfusion images from the model. These features were further used to train the parameters of artificial neural network model. Finally, the artificial neural network model was used to segment images. Glass and Reddick (1998) proposed a hybrid multi-scale back-propagation neural network to segment dynamic contrast-enhanced MRI images.

Compared with clustering method, machine learning method can learn the mapping relationship from feature space to training labels and improve the accuracy of tumor segmentation (Nowozin et al., 2010). However, this method still has some disadvantages. First, machine learning method needs the extraction of hand-craft features. The ability to describe features is the key to determining the accuracy of the classification results of model. Many of the methods for extracting hand-craft features only consider those regular edge information, texture information, etc. in natural images. They do not consider implicit characteristics of medical images, so they cannot obtain the effective information of medical images. Second, to improve the accuracy of classifier, a large number of features need to be computed to train the classifier. The computation task for feature extraction and selection is heavy and complicated (Tripathy et al., 2016; Zhang et al., 2016). Thus, machine learning method is time-consuming and can take up too much memory at runtime (Havaei et al., 2017).

The convolutional neural network (CNN)-based method has made great breakthrough in image recognition field (Bertasius et al., 2015; Chen et al., 2015a,b; Liu et al., 2015; Xie and Tu, 2015). Many researchers have applied this method in medical image segmentation (Rao et al., 2015; Roth et al., 2015; Zhang et al., 2015; Lyksborg et al., 2015). The CNN-based method can learn a hierarchy of increasingly complex features directly from patches (a local region of the image) (Pereira et al., 2016). Thus, there is no need to extract hand-craft features and the segmentation accuracy can be improved significantly.

Nevertheless, CNN-based method also has some disadvantages. This method operates on patches. Each pixel and pixels in its surrounding regions can form a patch. The rate of overlap between patches is quite high, so there exists high computational redundancy (Ronneberger et al., 2015). Clearly, this method is time-consuming and computationally expensive. Moreover, the size of patch limits the receptive field size of network. The network cannot focus on both the global and local features of images. In fact, almost all CNN-based methods based on patch need a post-processing procedure. Long et al. (2015) proposed a more straightforward network named fully convolutional network (FCN) for image segmentation. FCN is a neural network that is trained end-to-end. It can accept arbitrary-sized images as input and directly output segmentation images with the same size as the original image. In addition, it does not need any post-processing procedure. FCN has wide application in natural image segmentation and medical image segmentation. Ronneberger et al. (2015) constructed a U-Net architecture based on FCN and applied it in the segmentation of electron microscopy images. However, both FCN and U-Net have only one supervised output layer. With the network depth increased, there was not enough guidance of the network during the training. Thus, mis-segmentation often occurs for segmenting medical images.

Considering this disadvantage of FCN and U-Net, this paper proposed a multiple supervised residual network (MSRN) for the segmentation of osteosarcoma in CT images. This method added three supervised side output modules in residual network. Each side output module can compute the loss between probability map and ground truth at the scale of this module and back-propagate the loss value, guiding the multi-scale feature learning of network. Finally, the final segmentation results were obtained by fusing the results output by the three side output modules.

2. Materials and methods

2.1. Data

To test the segmentation performance of the MSRN algorithm, a small osteosarcoma dataset was built. This dataset included 2305 CT images of 23 osteosarcoma patients. Among them, 1900 CT images of 15 osteosarcoma patients were used as training data. The remaining 405 CT images of 8 osteosarcoma patients were used as testing data. All image data were from Beijing Jishuitan Hospital (Beijing, China).

The details of the image data were as follows. (1) The image data were obtained with enhanced CT. (2) The enhanced CT included GE CT, Philip CT and Toshiba CT equipment. The axial pixel resolution of image was 512 × 512. The image thickness obtained by the Philip CT is 5 mm and the sickness of the Toshiba CT and Philip CT is 1 mm. (3) For all the patients, osteosarcoma was in the leg.

An experienced radiologist (A) was asked to outline the osteosarcoma region in CT images. Considering that there may be error in the results of manual outlining, we randomly selected 100 out of the 2035 CT images and asked radiologists A and B to re-outline the osteosarcoma region. Statistical analysis indicated that the inter-observer intra-class correlation coefficient of the outlining results by radiologists A and B was 0.998 (95% confidence interval 0.997–0.999). This indicated that the outlining results by radiologists A and B were highly consistent. Besides, the intra-observer intra-class correlation coefficient of the two outlining results by radiologist A was 0.999 (95% confidence interval 0.998–0.999). This indicated that the two outlining results by radiologist A had no significant difference. Finally, the osteosarcoma regions outlined by radiologist A were used as the labels for network training and as the criteria for evaluating the segmentation results of MSRN.

2.2. Data pre-processing

To make the data more suitable for segmentation, we pre-processed the data.

(1) Outlining interested region in the image. Osteosarcoma was mainly located in the leg. However, there were many non-leg regions in the CT image. To reduce computation time, a 320 × 320 leg region was outlined as the interested region.
(2) Normalizing the CT image. In this experiment, CT images were from different equipment, which led to the different distribution of density in the same tissue in different CT images. Therefore, CT images were normalized and the formula was as follows (Jain and Bhandare, 2011).

$$I_{norm} = \frac{I - I_{\min}}{I_{\max} - I_{\min}} \tag{1}$$

where I was CT image for training, I_{norm} was normalized CT image, $I_{\min}$ was the minimum density of the image and $I_{\max}$ was the maximum density of the image.