Physica Medica 50 (2018) 13–19

Contents lists available at ScienceDirect

Physica Medica

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

Original paper

Fully automatic and robust segmentation of the clinical target volume for radiotherapy of breast cancer using big data and deep learning

Kuo Men, Tao Zhang, Xinyuan Chen, Bo Chen, Yu Tang, Shulian Wang, Yexiong Li*, Jianrong Dai*

National Cancer Center/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100021, China

ARTICLE INFO

Keywords:
Breast cancer radiotherapy
Automatic segmentation
Clinical target volume
Big data
Deep learning

ABSTRACT

Purpose: To train and evaluate a very deep dilated residual network (DD-ResNet) for fast and consistent auto-segmentation of the clinical target volume (CTV) for breast cancer (BC) radiotherapy with big data.

Methods: DD-ResNet was an end-to-end model enabling fast training and testing. We used big data comprising 800 patients who underwent breast-conserving therapy for evaluation. The CTV were validated by experienced radiation oncologists. We performed a fivefold cross-validation to test the performance of the model. The segmentation accuracy was quantified by the Dice similarity coefficient (DSC) and the Hausdorff distance (HD). The performance of the proposed model was evaluated against two different deep learning models: deep dilated convolutional neural network (DDCNN) and deep deconvolutional neural network (DDNN).

Results: Mean DSC values of DD-ResNet (0.91 and 0.91) were higher than the other two networks (DDCNN: 0.85 and 0.85; DDNN: 0.88 and 0.87) for both right-sided and left-sided BC. It also has smaller mean HD values of 10.5 mm and 10.7 mm compared with DDCNN (15.1 mm and 15.6 mm) and DDNN (13.5 mm and 14.1 mm). Mean segmentation time was 4 s, 21 s and 15 s per patient with DDCNN, DDNN and DD-ResNet, respectively. The DD-ResNet was also superior with regard to results in the literature.

Conclusions: The proposed method could segment the CTV accurately with acceptable time consumption. It was invariant to the body size and shape of patients and could improve the consistency of target delineation and streamline radiotherapy workflows.

1. Introduction

Radiotherapy of tumors requires accurate, patient-specific treatment planning to deliver high radiation doses to the target and to spare healthy tissues. Segmentation of the clinical target volume (CTV) and organs at risk (OARs) are essential steps for successful treatment delivery. In general, such segmentation is performed by manual delineation in computed tomography (CT) and/or magnetic resonance (MR) images. However, manual delineation is challenging, time-consuming, and subjective, with considerable inter- and intra-observer variability [1–4]. Thus, accurate automated segmentation methods are highly desired and useful for pre-treatment radiotherapy planning and adaptive radiotherapy during treatment.

A variety of automated segmentation software has been introduced to radiotherapy treatment planning. The common aspect of all these methods is that they are based on multiple-atlas [5–10] segmentation approaches. They use deformable image registration (DIR) methods to match the target image to be segmented and multiple atlases containing ground truth (GT) segmentations and then propagate the labeled structures in the atlas image onto the target image automatically. However, the DIR is not able to account for large deformation between source and target image. Therefore one “generic atlas” may not perform well for patients with considerable variation in the appearance of anatomic structures. Multiple atlases usually need to build according to the size, shape, or other inconsistencies in clinical image data. A potential solution is the feature-based machine learning approach, which could capture such variation and build into the prediction model. One “generic” model accounting for all variation will be more efficient in clinical practice.

In recent years, deep learning methods have resulted in many achievements in computer vision [11–19]. There has been increasing interest in applying deep learning to radiotherapy [20–26]. Ibragimov et al. [24] used typical convolutional neural networks (CNNs) for the segmentation of OARs for CT images of the head and neck. However, auto-segmentation of the CTV is more challenging for three main reasons. First, low contrast visibility and high noise levels usually lead to ambiguous and blurred boundaries between the CTV and normal tissues on CT images. Second, the CTV usually includes tissues with potential

* Corresponding author.
E-mail addresses: yexiong12@163.com (Y. Li), dai_jianrong@163.com (J. Dai).

https://doi.org/10.1016/j.ejmp.2018.05.006
Received 24 December 2017; Received in revised form 3 May 2018; Accepted 4 May 2018
1120-1797/ © 2018 Associazione Italiana di Fisica Medica. Published by Elsevier Ltd. All rights reserved.

tumor spread or subclinical diseases that are barely detectable in the planning CT images. Third, delineation of the CTV is highly dependent upon the physician's knowledge and recognition of structures. Hence, accurate segmentation of the CTV has become a "bottleneck" in radiotherapy. Recently, we developed deep dilated convolutional neural network (DDCNN) [25] and deep deconvolutional neural network (DDNN) [26] for automatic segmentation in rectal and nasopharyngeal cancer. Use of deep learning to carry out CTV segmentation for breast cancer (BC) has not been described before.

There are three main types of layers to build CNNs architectures: convolutional layer, pooling layer, and fully-connected layer [12]. These layers are stacked to form a full CNNs architecture. The levels of extracted features can be enriched by the number of stacked layers (depth). The network depth is of crucial importance, and the performance could greatly benefit from very deep models [27]. In this work, we trained and evaluated a much deeper network named deep dilated residual network (DD-ResNet) for automatic segmentation in planning CT of BC. We expected the accuracy could be improved with the deeper model stacked with more layers. The typical procedure for radiotherapy after breast-conserving surgery was studied. The proposed method was invariant to the body size and shape of patients with enough training examples. The method could "learn" such knowledge by itself and handle input images with huge differences. These are necessary preparatory and important steps towards developing automatic treatment planning methods for radiotherapy.

2. Materials and methods

2.1. Data acquisition

Data of patients with early-stage BC who underwent breast-conserving therapy from January 2013 to December 2016 in Department of radiation oncology, Cancer Hospital, Chinese Academy of Medical Sciences were collected. Patients received adjuvant radiotherapy after lumpectomy. Only those patients who received whole-breast radiotherapy were included in our study. Patients who underwent axillary or supraclavicular radiotherapy were excluded from the present study. The CTV included most ipsilateral breast tissue.

The data for the planning CT were acquired on Somatom Definition AS 40 (Siemens Healthcare, Forchheim, Germany) or Brilliance CT Big Bore (Philips Healthcare, Best, the Netherlands) systems set on helical scan mode. CT images were reconstructed using a matrix size of 512×512 and thickness of 5 mm. Delineation of the CTV was approved by senior radiation oncologists.

In total, 57,878 CT slices were collected from 800 patients. Four-hundred patients had right-sided BC and the remainder had left-sided BC. "Standard GT segmentations" were defined as the reference segmentations generated and cross-checked by experienced radiation oncologists. All the voxels that belonged to the GT segmentations of the CTV were extracted and labeled as the "outputs".

2.2. Deep learning algorithm for segmentation

Although the CTV includes the invisible tumor extent, deep learning is able to mine the non-explicit image properties hidden within in the image data. The data for model training is based on the CTV labels and their corresponding planning CT images. The physicians contoured the CTV on the planning CT images according to the breast-conserving surgery. The process of inferring CTV from the planning CT images after breast-conserving surgery is implicit in these data and has statistical regularity. As long as this statistical regularity exists, the deep learning method that is a data-driven statistical machine learning algorithm can learn the contouring process well.

Fig. 1 is a flowchart of the deep learning-based segmentation method. It was an end-to-end segmentation framework that could predict pixel-wise class labels in CT images. The training set (comprising CT images and manual segmentation labels) was used to adjust the parameters to train a good segmentation model. Then, the test set was used to assess the performance of the model.

Deep learning methods such as CNNs exploit three mechanisms (a local receptive field, weight sharing, and subsampling) that reduce the number of parameters that must be learned in a model drastically. Here, we introduced a robust deep learning algorithm (DD-ResNet) to segment the CTV for BC radiotherapy. Fig. 2 shows the detailed architectures of DD-ResNet. We deployed a 4-stream dilated convolutional module before using the ResNet-101[28] networks. The dilated convolutional module is able to efficiently extract original context information by introducing different dilated factors. By setting different dilated factors, the filter can achieve large receptive fields; thus can extract the multi-scale contextual feature. After that, the multi-scale feature maps are added to a specific feature number and feed forward to the ResNet-101 networks. The ResNet-101 has 101 weighted layers, and it is a fully convolutional network architecture. Similar to image classification, the ResNet-101 network mainly extracts low-level, middle-level and high-level visual features. The final extracted features are utilized to achieve pixel-level classification task. In addition, deep convolutional networks are difficult to optimize due to the vanishing gradients, and the vanishing feature is harmful for semantic segmentation task. The residual networks (ResNet) solved this problem by adding "shortcut connections" that were summed with the output of the convolutional layers. An example of the residual block is shown in Fig. 2b. It took a standard feed-forward convolutional network and added skipped connections that bypassed a few convolutional layers at a time. Each bypass gave rise to a residual block in which the convolutional layers predicted a residual that was added to the input tensor of the block. For a given input image, we used a multiple-path dilated convolutional network strategy to extract the multiple-scale feature map. A batch-normalized (BN) option was used after each convolutional layer. Then, an element-wise rectified-linear non-linearity (ReLU) max (0, x) was applied. Downsampling was carried out by conv3_1, conv4_1, and conv5_1 with a stride of 2. Thus, the size of original input image was reduced with a factor of 8. Therefore, the output of sum layer needs to be interpolation to the original size, and execute pixel-level classification. In our work, we use bilinear-interpolation to recover the original size. The proposed residual learning framework was easy to optimize and could gain accuracy from a considerably increased depth.

2.3. Experiments

In this work, training and evaluation of right- and left-sided BC were undertaken separately. We performed a fivefold cross-validation, where the dataset was randomly divided into five equal-sized subsets. Firstly, we trained the model on the first 4 subsets (80% of the data) and tested on the 5th subset (20% of the data). Subsequently, we chose another subset as the test set and trained a second model on the remaining 4 subsets. We repeated this step until we trained 5 models. We implemented the training, evaluation, error analysis and visualization pipeline of our model using Caffe [29] (a publicly available deep learning framework) and then compiled with the NVIDIA CUDA® Deep Neural Network library (cuDNN) [30] computational kernels.

2.3.1. Training

We trained models for right- and left-sided BC separately. The training set was used to "tune" the parameters of the networks. In detail, the original two-dimensional (2D) CT images were the inputs and the corresponding segmentation probability maps about the CTV were the outputs. The model parameters for each network were initialized using the weights from the corresponding model trained on ImageNet and were then "fine-tuned" using BC data. We adopted data-augmentation methods such as "random cropping" and "flipping" to reduce overfitting. We used a batch size of 1 for DD-ResNet due to memory

Download English Version:

https://daneshyari.com/en/article/8248545

Download Persian Version:

https://daneshyari.com/article/8248545

Daneshyari.com