Supplementary files

Title

Faster Region-based Convolutional Neural Network-Aided Diagnosis of MRI Images of the Circumferential Resection Margin in Rectal Cancer: A Feasibility study

Authors

Dongsheng Wang, Jihua Xu, Zhengdong Zhang, Shuai Li, Xianxiang Zhang, Yunpeng Zhou, Xunying Zhang, Yun Lu*

Supplementary Methods

Faster RCNN principles and training processes

The automatic detection of the invasion of CRM in rectal cancer by Faster RCNN was studied. The method included a region proposal network (RPN) and Faster RCNN. The structure of the RPN network was to add a convolution layer to the last layer of the existing network structure, such as VGG and ZF, which could be used to distinguish all of the possible candidate frames on the extracted feature image. The RPN network performed feature extraction using a sliding-window strategy on the new convolutional layer, which overcame the time consuming and storage redundancy problems caused by extracting features on the original image using the sliding-window strategy. We first mapped a point on the new convolutional layer back to its corresponding point on the original image through a spatial pyramid and identified positive and negative labels at each position on the feature image by designing k different scales. The convolution layer had 4k outputs and encoded the coordinates of k bounding boxes. The classification layer outputted 2k scores and estimated the probability of whether each object was a target or not. By default, we used three scales and three length-width ratios and produced k=9 anchor points at each sliding position. To select the candidate area, we assigned a binary label (either a target or not) to each anchor point. We assigned a positive label to two types of anchor points: (1) an anchor point that overlapped the actual bounding box with the highest intersection over union (IoU) or (2) an anchor point that overlapped the actual bounding box more than 0.7 IoU. For all actual bounding boxes, if the IoU ratio of an anchor point was below 0.3, we assigned a negative label to the nonfrontal anchor point. In this way, we could generate a region on the convolutional feature image that might be the positive CRM, and then, we merged the adjacent regions to decrease the training candidate regions using the nonmaximum suppression method. Therefore, this approach reduced a large number of unnecessary repetitive calculations needed for subsequent target detection and classification. RPN and Faster RCNN shared the convolutional feature image, and through pooling the layer-of-interest area and following with two sub-fully connected layers, probability scores that could be used to predict the coordinates and categories of the bounding boxes were obtained. In the experiment, RPN and Faster RCNN were alternately trained in two stages, and the parameters were fine-tuned in the iterations. The position of the candidate box was calibrated by the bounding-box regression in such a way that the optimal result could be obtained finally.

During the training process, 1020 images identified as CRM-positive from 192 patients in the training group were used as the training data set, and the pretrained VGG16 with 13 convolutional layers and 3 fully connected layers in ImageNet was used for the initialization of the feature extraction network. The weights of the RPN and the feature vector Faster RCNN of the interesting region were assigned with random values that corresponded to a Gaussian distribution with a mean of zero and a deviation value. The training process was divided into two stages, which included 80,000 times training of the RPN candidate regions (the learning rate of the first 60,000 was 0.0001, and the learning rate of the last 20,000 was 0.00001) and 40,000 times classification and regression of the feature vectors of the candidate regions (the learning rate of the first 30,000 was 0.0001, and the learning rate of the last 10000 was 0.00001). The momentum was 0.9, and the weighted delay was 0.0005. The scales of an anchor of the RPN were set to 1282, 2562, and 5122, and the aspect ratios of the anchor were set to 0.5, 1, and 2. During the training, end-to-end back-propagation data were acquired using the stochastic gradient descent (SGD) method. The parameters of the deep learning network, such as the weights and other variables, were adjusted to reduce the loss function and have network convergence. The curve of the attenuation function of the Faster RCNN network is shown in Supplementary Figure 1.

Supplementary Figure 1: The loss function values of the entire network during the training process