Supplementary Materials for "Preoperative diagnosis of hepatocellular carcinoma patients with bile duct tumor thrombus using deep learning method"

Appendix A Model configurations

Main configurations of the model include the initial sizes and aspect ratios of the anchors, the intersection over union (IoU) threshold and the non-maximum suppression (NMS) threshold, and the parameters for image standardization. In order to determine adequate size and aspect ratio of the anchors, statistical analysis was carried out based on the annotations, including plotting histogram and calculating quantiles. From Figure A.1 and Table A.1, the aspect ratios of the bounding boxes concentrate between 0.50 and 3.50. Considering computational efficiency, the initial aspect ratios of the anchors were set to five values: 0.33, 0.50, 1.00, 2.00, 3.00. From Table A.2, the minimum, 25% quantile, median, 75% quantile, and maximum of the bounding boxes' sizes are approximately $16^2, 32^2, 48^2, 68^2, 168^2$ pixels, respectively. Therefore, the initial anchor sizes were set according to these quantiles with each size corresponding to one feature map output by FPN.

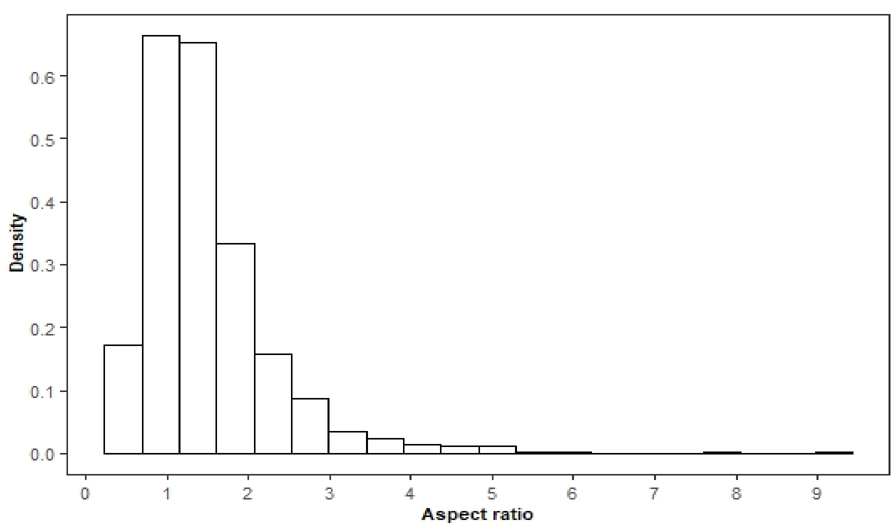

Fig. A.1 Histogram of aspect ratios of bounding boxes

Table A.1 Quantiles of aspect ratios of the bounding box

Prob	0%	5%	25%	50%	75%	95%	100%
Quantile	0.35	0.59	0.96	1.26	1.80	2.92	9.09

Table A.2 Quantiles of sizes of the bounding boxes

Prob	0%	25%	50%	75%	100%
Quantile	16.00^2	32.50^2	48.74^2	68.56^2	167.57^2

The IoU value is defined as the area of the overlapping part of the bounding box generated by the model and the ground truth box labeled by the doctor divided by the total area of the two boxes. IoU threshold is used to determine if an anchor is foreground or background. The Non-Maximum Suppression (NMS) threshold is used to filter out anchors with high degree of overlap, i.e. the redundant bounding boxes with higher overlap than NMS threshold will be removed when outputting detection results. The IoU threshold of the foreground was set to 0.70 for RPN so that the criteria for determining whether an anchor is positive is strict, and the NMS thresholds was set to 0.60 for both the RPN and R-CNN to reduce the number of anchors and improve training efficiency.

For Faster R-CNN, the input image should be standardized then scaled to a fixed size before extracting features. The images were normalized so that the sample mean and the standard deviation were 0 and 1, respectively. Then, the image were magnified four times its original size (twice the x-axis and twice the y-axis), so that the DBD could be enlarged without excessive deformation.

Appendix B Training details

Due to the limited data, data augmentation was conducted in the training process. There are many approaches to implementing data augmentation. Due to the tiny DBD size and the grayscale characteristic of CT images, some methods (eg., changing brightness and adding noise) are not suitable in the current study. During our training process, four techniques for data augmentation were used by turns: random horizontal and vertical translation, random cropping, random clockwise or counter-clockwise rotation up to 10 degrees, and random horizontal or vertical flips with a probability equal to 0.4, 0.4, 0.4, and 0.5.

In the model training, the stochastic gradient descent (SGD) algorithm was used as the optimizer and the L2 penalty coefficient (also known as weight decay) was set to 0.0005. The learning rate was set following a cosine annealing schedule with the initial learning rate being 0.0003. The period for the cosine annealing schedule was set to 40 epochs with the changing pattern of the learning rate being shown in Figure B.1. The Faster R-CNN models were trained for 120 epochs for each fold with a batch size of two. The model was trained on one graphic card (GTX 2080Ti; NVIDIA, Santa Clara, Calif) using Pytorch.

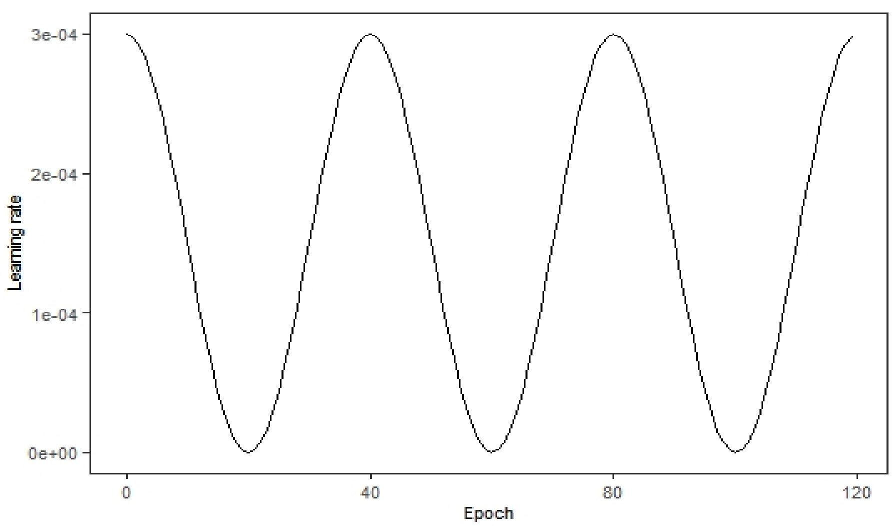

Fig. B.1 Change pattern of the learning rate