



## MRI Liver Image Assisted Diagnosis Based on Improved Faster R-CNN

Minjie Tao<sup>1</sup>, Jianshe Lou<sup>2</sup>, Li Wang<sup>1\*</sup>

<sup>1</sup> Department of Hepatobiliary Surgery, Affiliated Hangzhou First People's Hospital, Zhejiang University School of Medicine, Hangzhou 310006, China

<sup>2</sup> Audit Office of Zhejiang Business College, Hangzhou 310053, China

Corresponding Author Email: [13646846188@139.com](mailto:13646846188@139.com)

<https://doi.org/10.18280/ts.390428>

### ABSTRACT

**Received:** 5 May 2022

**Accepted:** 18 July 2022

#### Keywords:

*MRI images, liver occupancy, image segmentation, deep learning, Faster R-CNN*

In response to challenges in liver occupancy such a variety of types and manifestations and difficulties in differentiating benign and malignant ones, this paper takes liver images of enhanced MRI scan as the research object, targets on the detection and identification of liver occupancy lesion areas and determining if it is benign or malignant. Accordingly, the paper proposes an auxiliary diagnosis method for liver image combining deep learning and MRI medical imaging. The first step is to establish a reusable standard dataset for MRI liver occupancy detection by pre-processing, image denoising, lesion annotation and data augmentation. Then it improves the classical region-based convolutional neural network (R-CNN) algorithm Faster R-CNN by incorporating CondenseNet feature extraction network, custom-designed anchor size and transfer learning pre-training. This is to further improve the detection accuracy and benign and malignant classification performance of liver occupancy. Experiments show that the improved model algorithm can effectively identify and localise liver occupancies in MRI images, and achieves a mean average precision (mAP) of 0.848 and an Area Under the Curve (AUC) of 0.926 on the MRI standard dataset. This study has important research significance and application value for reducing manual misses and misdiagnosis and improving the early clinical diagnosis rate of liver cancer.

## 1. INTRODUCTION

Liver cancer is a common malignant tumour in clinical practice, with the 5th highest mortality rate among malignant tumours worldwide. China has about half of the world's liver cancer cases, with a significantly higher incidence and mortality rate than the world level, with the 4th highest incidence rate and the 3rd highest cause of cancer-related deaths [1-3]. Medical data of hepatobiliary surgery indicates that liver cancer is often found to be progressive or mid to late stage once detected, and the effective treatments that can be carried out are limited. At present, the radical cure with definite efficacy are mainly liver resection and transplantation, etc. Although surgery can effectively control the cancer development, reports [4-6] say that the five-year recurrence rate of patients after radical resection is as high as 40%-70%, resulting in a prognosis below desired level. Other studies have shown that the five-year survival rate for people with small hepatocellular carcinoma (a single nodule less than 3 cm in diameter) can reach 80% after surgical resection and radiofrequency ablation [7-10]. Therefore, early detection of small intrahepatic occupancies has great significance for the treatment and prognosis of liver cancer.

Among the liver occupancy diagnostic techniques, dynamic contrast-enhanced CT or MRI examinations are recommended by the American Association for the Study of Liver Diseases (AASLD) for the non-invasive diagnosis of hepatocellular carcinoma (HCC), and the American College of Radiology updated the Liver Imaging Reporting and Data System (LI-RADS v2018) in 2018, which aims to unify the imaging signs

and imaging diagnostic process for HCC. For the detection of earlier stage liver cancer, such as the previously mentioned small hepatocellular carcinoma (small HCC), a medical expert [11] compared the diagnostic value of CT and MRI imaging respectively based on LI-RADS v2018 for HCC that is less than 3 cm in diameter, and demonstrated that MRI imaging was overall superior to CT examination for diagnosis of both obvious cancer-caused liver occupancy and early small HCC, MRI imaging has higher enhanced sensitivity and imaging saliency. In view of this, to improve the diagnostic rate of liver occupancies, this paper uses MRI-enhanced scans to study the data.

With the advent of big data in healthcare, there is an urgent need for computer-aided diagnosis (CAD) techniques that can quantitatively analyse medical images and give proactive references. In recent years, deep learning algorithms (e.g., Fast R-CNN, Deep Convolutional Neural Network and Faster R-CNN) have achieved good results in liver tumour detection and recognition. Meng et al. [12] proposed a 3D dual path multi-scale convolutional neural network that used pairwise paths to balance the performance of segmentation and reduce the computational resource requirements for robust segmentation of the liver and liver tumours. Tang et al. [13] used Faster R-CNN to detect the approximate location of the liver and then fed into DeepLab to segment the liver. Li et al. [14] proposed a Hybrid-DenseNet (H-DenseNet), which effectively aggregated the 2D DenseUNet extracted intra-slice features into a 3D DenseUNet to perform 3D segmentation of the liver and tumour simultaneously. Bousabarah et al. [15] used a deep convolutional neural network with radiological

capabilities to automatically detect and characterise hepatocellular carcinoma on contrast-enhanced MRI. Kim et al. [16] used a deep learning-based classifier to detect HCC on contrast-enhanced MRI. Zhao et al. [17] combined adversarial learning ideas with Fast R-CNN to improve the detection capability of the network using the three-way adversarial idea. While the above studies have demonstrated the feasibility of deep learning techniques for tumour target detection, there are fewer studies related to image-aided diagnosis for the lesion classification of liver occupancy and determination of benign and malignant liver occupancies. In addition, the problems of uneven intensity, noise interference, weak contrast and irregular appearance and size of tumour lesions in MRI [18] pose challenges for CAD research of liver occupancy images based on deep learning techniques.

To address the above issues, this paper proposes an auxiliary diagnosis algorithm for detecting lesion types and identifying benign and malignant liver occupancies using image segmentation and deep learning techniques. The paper demonstrates the efficacy of this technique through experiments in predicting benign and malignant liver lesions, clarifies the efficacy of this improved algorithm in distinguishing different categories of liver occupancies after confirming liver lesions, and explores the feasibility and application value of this improved algorithm for the identification and detection of liver lesions in MRI liver images, so as to assist physicians in analysing MRI images of liver cancer and making further diagnostic measures.

## 2. CONSTRUCTION OF A STANDARD DATASET FOR MRI-BASED LIVER OCCUPANCY DETECTION

Deep learning and CAD algorithms require a large amount

of training data, but there is a lack of standard MRI image datasets for liver occupancy. In response, we decided to construct a standard MRI dataset for liver occupancy detection in this paper, including both benign and malignant occupancies, for training and testing of deep learning algorithms. The images that dynamic contrast-enhanced MRI (DCE-MRI) acquire are mostly multimodal data of the liver depending on the time of contrast agent injection [19], including iso-inverted phase T1WI, pressurised lipid/non-pressurised lipid T2WI, unenhanced scan, diffusion-weighted imaging (DWI) and enhanced scan sequence images (arterial phase, portal phase, equilibrium phase, hepatobiliary phase). A hepatocyte-specific contrast agent is used during the hepatobiliary phase.

### 2.1 Image pre-processing

The main tasks during the data pre-processing stage were: (1) the raw MRI data in DICOM format acquired from the hospital PACS system were converted to JPEG images that could be processed for deep learning analysis by using Matlab R2018a to transcode the DICOM files [20]; (2) by collaborating with physicians with years of radiology experience at the partner hospital, we analysed the JPEG images of liver occupancy patients admitted to the hospital in the last year, and included 93 liver occupancy patients in the cases who met the following two criteria: (1) DCE-MRI performed within seven days before biopsy or treatment; (2) patients with a diagnosis of benign or malignant liver occupancy confirmed by surgery or puncture biopsy. A total of 93 liver occupancy patients aged 30 to 80, including 35 women and 58 men, were finally included in this study (Table 1).

**Table 1.** The current main MRI benign and malignant liver occupancy types and MRI signs

Occupancy type		Major MRI signs
Benign liver occupancy	Hepatic hemangioma	Prevalent in women aged 40-50. MRI shows moderate to high signal on T2WI and low signal on T1WI. Enhanced scans show nodular discontinuous enhancement and contrast agent retention at the edges. Central necrosis is seen in large hepatic haemangiomas. If hepatocyte-specific contrast agent is used, there may be an artifact of contrast agent outflow.
	Focal nodular hyperplasia (FNH)	Prevalent in young women. MRI shows isosignal T1WI and mild high-signal T2WI. The central necrosis may have low signal on T1WI and moderate to high signal on T2WI. On enhancement, the arterial phase is homogeneous, the portal phase is isosignal to the liver parenchyma, and the central necrosis is seen as delayed enhancement. FNH shows no rapid contrast washout.
	Hepatic adenoma (HCA)	Prevalent in patients using oral estrogen. MRI shows mild/moderate high signal on T2WI with arterial-phase intensification. Pathologically, HCA is classified into the following three types with different imaging features: 1. Inflammatory type, with marked high signal at the margins on T2WI and delayed enhancement on enhancement scans. 2. HNF-1 $\alpha$ -activated HCA with diffuse fatty component, i.e., high signal on T1WI and antiphase signal decrease. 3. Beta-chain protein-activated type, with indistinct irregular margins and high signal on T2WI. This type tends to malignant transformation. Note: HCA is sometimes not easily distinguished from FNH, but there is usually no necrosis within an HCA. The use of a hepatocyte-specific contrast agent can help to differentiate them. FNH shows contrast uptake, but generally HCA does not.
	Cystic lesions	The cysts are usually benign. MRI shows uniform low signal on T1WI and significant high signal on T2WI, with clear margins and no intensification after enhancement.
Malignant liver occupancy	Hepatocellular carcinoma (HCC)	The main signs include "envelope", significant enhancement in the non-circular arterial phase, and non-circular "contouring". MRI shows high signal in the arterial phase, contrast agent outflow in the portal phase, low signal in T1WI, mildly high signal in T2WI and high signal in DWI. The signal is heterogeneous in the early enhanced arterial phase, with contrast agent outflow and pseudo-envelope patterns seen in the late enhanced phase.
	Intrahepatic bile duct cancer	MRI shows low signal on T1WI and high signal on T2WI, with heterogeneous continuous enhancement at the edges and retraction of the hepatic tegument.
	Metastatic cancer of the liver	MRI presentation shows multiple lesions of variable size with low signal on T1WI and high signal on T2WI. Most of the metastases have rich blood supply, with circumferential enhancement seen in segment VII lesions and contrast agent outflow in lesion's late enhancement.

## 2.2 Image denoising

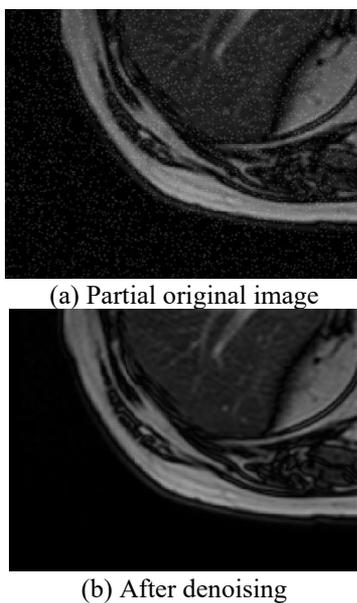
For building a well-performing CAD system, it is essential to improve the image quality of DCE-MRI through reasonable image denoising. What MRI images generate is mainly thermal noise and sometimes physiological noise [21]. Many studies have suggested that it belongs to Rician noise [22], which is strongly correlated with signal [23]. Traditional denoising methods are only suitable for filtering certain types of noise, but not for filtering Rician noise, while wavelet transform has better filtering effect on Rician noise. Therefore, this paper chooses a wavelet transform-based denoising method [24] to denoise the image samples. The main process has three steps: firstly, we input the original MRI liver image with noise and added additive Gaussian noise to the original signal in the data; secondly, we performed wavelet change to obtain the wavelet coefficient matrix; and finally, the matrix was processed by hard and soft thresholding functions. After setting a threshold, we reduced and zeroed coefficients larger and smaller than  $\phi$  respectively. Then, we obtained the denoised image based on the new coefficients. The soft and hard thresholding methods add absolute value judgement to the above, and the equations are expressed as follows:

$$\rho(\psi) = \begin{cases} \text{sign}(\psi_{i,j}) \cdot (|\psi_{i,j}| - \phi), & |\psi_{i,j}| \geq \phi \\ 0, & |\psi_{i,j}| < \phi \end{cases}$$

$$\rho(\psi) = \begin{cases} \psi_{i,j}, & |\psi_{i,j}| \geq \phi \\ 0, & |\psi_{i,j}| < \phi \end{cases}$$

where,  $i$  is the number of decomposition layers and  $j$  represents the wavelet coefficients in different directions. After soft and hard thresholding methods for wavelet denoising, we found that the images obtained by the soft thresholding method were smoother, while the image texture by the hard thresholding method had more visible jitters, hence we finally chose the denoised images obtained through soft thresholding.

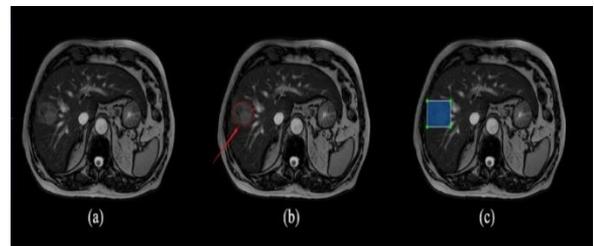
Figure 1 below shows the denoising effect of this research (some areas are presented in enlargement).



**Figure 1.** Denoising effect of MRI images through soft thresholding

## 2.3 Lesion labeling

The annotation of the dataset in this study was performed under the guidance of a specialist radiologist at the partner hospital. We determined the location of the liver occupancy and the nature of the case after taking account of the patient's diagnostic history. The liver cancer in this dataset was classified as benign or malignant. The lesion location was annotated using the minimum coverage matrix that could completely cover the lesion, using the target detection annotation software LabelImg [25]. After manual annotation of the images, the software converts the annotation information into an XML format file for storage, which is flexible enough to store the location and category-structured data of the masses for the deep learning algorithm to read during training. As shown in Figure 2, (a) is the original image, (b) is the physician's manual annotation of the lesion location, and (c) is the software representation in terms of a minimum coverage matrix.



(a) Original MRI image (b) Annotation of lesion location by physician (c) Annotation by software

**Figure 2.** Image annotation

## 2.4 Image data augmentation

In deep learning, an adequate number of samples is required to ensure the effectiveness of the training model and the generalization ability of the model [26]. Therefore, to obtain sufficient training data, this study increases the data volume of the dataset by means of data augmentation, in the expectation that the image texture and pathological features of the limited base images will be expressed in the augmented images to increase the sample space. In this paper, we mainly adopted the geometric transformation of image rotation and flip for data augmentation, for which we rotated each image counterclockwise by  $60^\circ$ ,  $90^\circ$ ,  $180^\circ$ , and  $270^\circ$  as well as horizontal and vertical flips. Here, rotation of an image means that each pixel point is rotated by an equal angle at the same origin. Its affine transformation formula is

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} \cos \theta & \sin \theta & 0 \\ -\sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_0 \\ y_0 \\ 1 \end{bmatrix}$$

Calculation formula for coordinates after horizontal flip:

$$\begin{bmatrix} x_1 & y_1 & 1 \end{bmatrix} = \begin{bmatrix} x_0 & y_0 & 1 \end{bmatrix} \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ width & 0 & 1 \end{bmatrix}$$

$$= [width - x_0 \quad y_0 \quad 1]$$

Calculation formula for coordinates after vertical flip:

$$\begin{aligned}
 [x_1 \ y_1 \ 1] &= [x_0 \ y_0 \ 1] \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & \text{height} & 1 \end{bmatrix} \\
 &= [x_0 \ \text{height} - y_0 \ 1]
 \end{aligned}$$

After counterclockwise rotation in the four angles and flips in the two directions, the original, rotated and flipped MRI

images of the liver formed the training data for subsequent deep learning. Figure 3 below shows the sequence map of one case's T2 image after image augmentation: (a) is the original image, (b) is the image after rotating 60 degrees counterclockwise, (c) is the one after rotating 90 degrees counterclockwise, (d) is the one after rotating 180 degrees counterclockwise, (e) is the one after rotating 270 degrees counterclockwise, (f) is the image after horizontal flip, and (g) is the one after vertical flip.

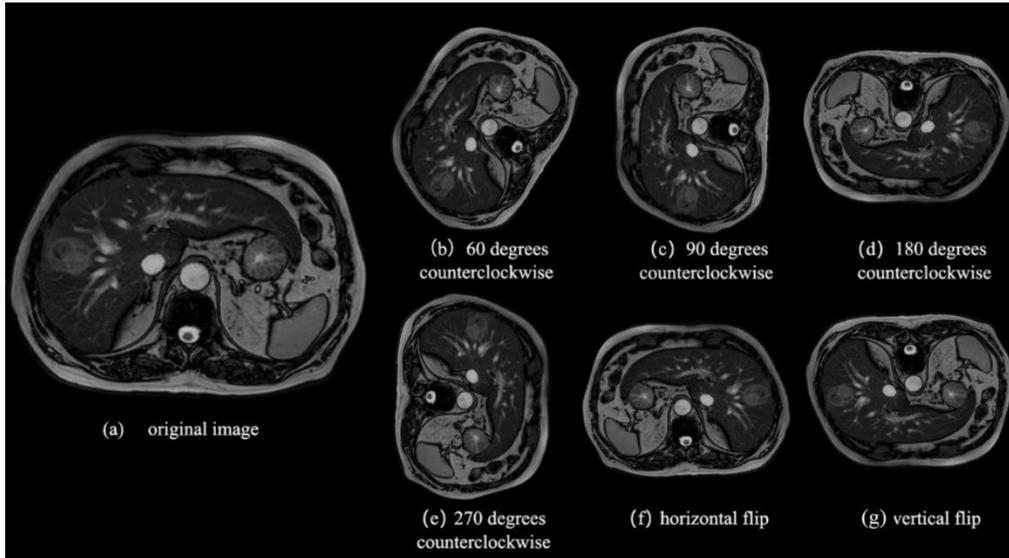


Figure 3. MRI image data augmentation

### 3. UNET FOR LIVER SEGMENTATION

MRI images are scanned in the abdominal region, so each MRI image of the data covers four organs—spleen, liver, left kidney and right kidney. Our goal is to segment the liver area to facilitate subsequent classification and identification of the target of interest. Segmentation of the liver area is a prerequisite for subsequent feature extraction and accurate classification, as well as an important step in the quantitative analysis of tumours by physicians.

UNet is widely used in medical image segmentation [27]. Its advantages are: (1) multi-scale information extraction: both details and coarser abstract information are effectively extracted and retained; the gradient information of fuzzy boundaries is maximally retained while reducing the impact of noise. (2) skip connection: the more accurate information on gradient, point and line of the encoder at the same layer is directly concatenated into the decoder at the same layer, which is equivalent to adding detailed information to the general area of the target to make UNet obtain more accurate segmentation results. Considering that the UNet structure is not only smaller in model but also higher in accuracy [28], this paper uses UNet to finish liver segmentation from abdominal MRI data with the following network architecture diagram:

In Figure 4, the left side is repeated downsampling->convolution, and the right side is repeated upsampling->convolution. The first part of the figure is feature extraction, where there is a scale for each passing through a pooling layer. In the upsampling part, each upsampling is fused with the same scale of the channel corresponding to the feature extraction part (labelled copy and crop in the figure). But it is cropped before the fusion. The fusion here is stitching.

The blue arrow represents a 3x3 convolution operation with a stride of 1 and valid padding, so that after each convolution, the feature map size is doubled. The red arrows represent a 2x2 max pooling operation. Since the 2\*2 max-pooling operator is suitable for images of even pixel length and width, it is important to choose the right input size. The green arrows represent a 2x2 convolution plus upsampling operation, which multiplies the feature map size by 2. The grey arrows represent a copy and cut operation, where it can be noticed that the last layer on the left side of the same layer has a slightly larger resolution than the first layer on the right side, which leads to some cutting if you want to make use of the features in the shallower layers. The last layer of the output is classified using a 1x1 convolutional layer, and the two layers of the output are foreground and background. A comparison of the experimental segmentation results and physicians' annotation result in this paper is shown in Figure 5.

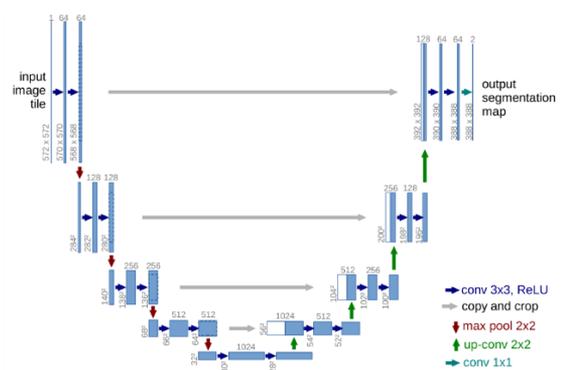
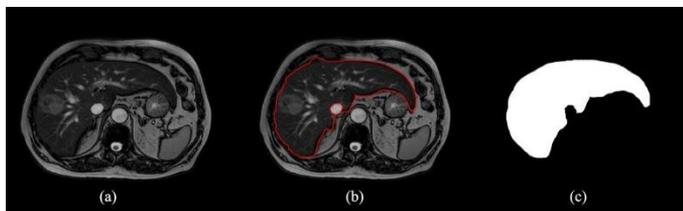


Figure 4. UNet architecture



(a) Original image (b) segmentation result annotated by physicians (c) Experimental segmentation result

Figure 5. Results of liver segmentation

#### 4. MRI LIVER IMAGE ASSISTED DIAGNOSIS BASED ON IMPROVED FASTER R-CNN

Target detection of liver MRI images refers to the localisation and diagnosis of liver occupancy targets from MRI image data. Accurate localisation of liver occupancy is the fundamental basis for assisting physicians in surgical planning, interventional surgery, and tumour definition. The detection results are integrated with patient age, clinical comorbidities, and biochemical results for guiding the post-operative treatment of patients. In terms of deep learning target detection, the R-CNN algorithm Faster R-CNN integrates image feature extraction, pre-selected box extraction, target regression, and target classification into a single network to achieve an efficient and unified end-to-end target detection algorithm [29]. It delivers superior performance in terms of both detection speed and accuracy. Therefore, this paper uses the Faster R-CNN framework as the basis, and proposes an improved Faster R-CNN algorithm for the integration of recognising and classifying benign and malignant occupancy from MRI liver images, given the multiple types of liver occupancies with complex and varied size and morphology.

##### 4.1 Network model design

There are 3 steps for Faster R-CNN to detect the input liver segmentation images: firstly, the target features in the input images are extracted by the pre-trained feature extraction network; then the region proposal network (RPN) uses the extracted features to find a certain number of regions of interest (ROI) to estimate the class and location of the target

that may contain the lesion. The image features and ROI are input to the ROI pooling unit in the Faster R-CNN to extract features, and Softmax regression is used to classify the ROIs and determine the class of liver occupancy, while fine-tuning the positions of these ROIs using bounding box regression to obtain the final accurate position of the detection box, i.e., to localize the lesion. The network architecture of the Faster RCNN is as follows Figure 6 shows.

##### 4.2 CondenseNet feature extraction network

Some scholars have demonstrated that for small target detection in medicine, if DenseNet is employed as a feature extraction network for Faster R-CNN, the experimental performance outperforms the VGG16 as well as the ResNet structure employed in the original Faster R-CNN [30].

However, one of DenseNet's biggest drawbacks is the large video memory consumption, mainly due to the generation of more extra feature layers. To reduce the memory consumption of the model during training, Gao Huang at Cornell University [31] optimized the DenseNet network in 2018 by using convolutional group operations and pruning during training to reduce memory and increase speed, making it more computationally efficient and storing fewer parameters. Hasan and Linte [32] used CondenseUNet in 2020 for biventricular blood pooling and myocardial segmentation in cardiac cine MRI (CMR) imaging. Experiments demonstrated that the CondenseUNet architecture can be used in the Automated Cardiac Diagnostic Challenge (ACDC) dataset, using half (50%) of the memory requirements of DenseNet and one-twelfth (approximately 8%) of the memory requirements of UNet, while still maintaining excellent cardiac segmentation accuracy. Accordingly, this study uses the CondenseNet network architecture for feature extraction of the dataset to obtain better network performance, while being suitable for MRI images and ensuring memory requirements. The CondenseNet network architecture is characterised by: (1) the introduction of convolutional group operations, with an improvement during the introduction of group operations in 1\*1 convolution. (2) Pruning of weights is done at the beginning of training, instead of pruning the trained model. (3) Introducing cross-block dense connectivity on top of DenseNet. The CondenseNet network configuration used in this dataset is shown in Table 2.

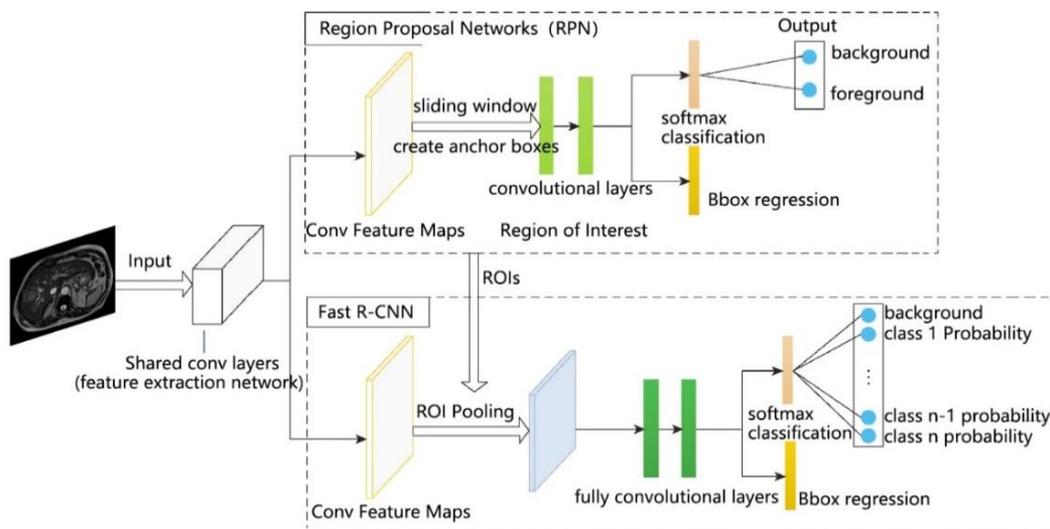


Figure 6. Faster R-CNN architecture

**Table 2.** CondenseNet network architecture table

Feature map size	structure
112 × 112	3 × 3 Conv, stride 2
112 × 112	$\begin{bmatrix} 1 \times 1 & L\text{-conv} \\ 3 \times 3 & G\text{-conv} \end{bmatrix} \times 4 \quad (k = 8)$
56 × 56	2 × 2 average pool, stride 2
56 × 56	$\begin{bmatrix} 1 \times 1 & L\text{-conv} \\ 3 \times 3 & G\text{-conv} \end{bmatrix} \times 6 \quad (k = 16)$
28 × 28	2 × 2 average pool, stride 2
28 × 28	$\begin{bmatrix} 1 \times 1 & L\text{-conv} \\ 3 \times 3 & G\text{-conv} \end{bmatrix} \times 8 \quad (k = 32)$
14 × 14	2 × 2 average pool, stride 2
14 × 14	$\begin{bmatrix} 1 \times 1 & L\text{-conv} \\ 3 \times 3 & G\text{-conv} \end{bmatrix} \times 10 \quad (k = 64)$
7 × 7	2 × 2 average pool, stride 2
7 × 7	$\begin{bmatrix} 1 \times 1 & L\text{-conv} \\ 3 \times 3 & G\text{-conv} \end{bmatrix} \times 8 \quad (k = 128)$
1 × 1	7 × 7 global average pool 1000D fully-connected, softmax

### 4.3 RPN and anchor design

The function of RPN is to generate candidate regions for liver occupancy detection. For any feature map received, RPN can compute a series of candidate regions and a corresponding score between 0 and 1, indicating the confidence level that the candidate region is predicted to be a foreground target. To generate candidate regions, RPN uses a sliding window of size 3×3 to obtain  $n \times n$  anchor locations based on the shared feature map. RPN also uses  $k$  different shapes of anchors in the process to enrich the prediction range for each sliding window location. One anchor position yields  $k$  candidate regions, so that for an input feature map of size  $W \times H$ , RPN obtains  $W \times H \times k$  anchors with translation invariance. In this study, the size of the lesion occupancy was statistically analysed for 93 patients in the constructed dataset, and the size of the occupancy ranged from 8mm to 80mm. Based on the proportion of this statistic on the original MRI image and the corresponding perceptual field size of the shared convolutional feature map mapped to the CondenseNet output, we designed three different scales and three different aspect ratios, which were combined into nine different shapes of anchors, namely  $72^2$ ,  $288^2$  and  $512^2$ , with aspect ratios of 1:1, 2:1 and 1:2 respectively.

### 4.4 Transfer learning model training

To obtain good prediction performance, we also had to employ transfer learning to train the network model while addressing the problem of insufficient data volume. This is because although we have performed data augmentation, the data volume of the MRI liver target detection dataset we constructed is still relatively too small compared to the number of neural network parameters, which tends to cause overfitting of the network parameters during training and poor recognition results. Therefore, we need to pre-train the improved network model on a natural image open dataset with a large data volume beforehand, so that the network learns certain natural image texture patterns in advance to obtain the model parameters to initialize our model, and then fine-tune

the network on the liver occupancy dataset afterwards.

The commonly used open datasets for natural images include ImageNet, an image classification dataset, and PascalVOC, a target detection dataset. ImageNet contains more than 1.5 million annotated natural images covering over 1000 item categories. Pascal VOC consists of Pascal VOC 2007 and Pascal VOC 2012, together containing a total of more than 30,000 images, 70,000 detection targets and 20 categories [33]. Transfer learning can be divided into partial transfer learning and full transfer learning. Partial transfer learning refers to loading some of the network architecture parameters from a pre-trained model, such as loading only a few specific convolutional layers; full transfer learning refers to loading the complete network parameters from the pre-trained model. In the Faster R-CNN training in this study, we used both.

(1) Pre-training CondenseNet on ImageNet by first performing partial transfer learning of the feature extraction network.

(2) Full transfer learning was then performed, with the Faster R-CNN structure pre-trained on the Pascal VOC 2007+2012 dataset and finally fine-tuned on the MRI liver occupancy dataset based on the obtained network parameters.

## 5. EXPERIMENTAL RESULTS

To demonstrate the effectiveness of the Faster R-CNN optimisation and improvement in this paper, two evaluation metrics are used to assess the detection and classification performance of the Faster R-CNN model trained in this paper. The first evaluation metric is the Mean Average Precision (mAP), which is commonly used in target detection, and the other is the Free-response Receiver Operating Characteristic (FROC) curve. A comparison was made between the detection and classification performance of a model trained using the original Faster R-CNN network and the improved model in this paper. The original Faster R-CNN refers to the model obtained by using VGG16 as the backbone network and trained based on the original anchor size and without using transfer learning.

The experimental evaluation was performed on the MRI liver dataset constructed in this paper, which was derived from the MRI-enhanced liver scans of 93 patients, among whom 15 were with benign occupancies and 78 with malignant occupancies. The dataset contained a total of 3,906 original MRI liver images and data augmented images, of which 558 were original MRI liver images. Among the 558 images, 90 were benign occupancies and 468 were malignant occupancies.

### 5.1 Mean average accuracy

The accuracy of the liver occupancy detection algorithm for a given liver occupancy category  $A$  is calculated by the following formula:

$$precision_A = \frac{TP}{TP + FP} = \frac{N(TruePositives)_A}{N(GroundTruths)_A}$$

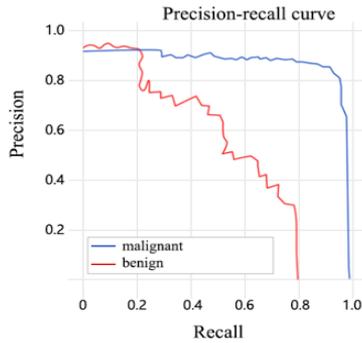
The average accuracy AP value refers to that under the assumption that each MRI image on the test set contains true annotations for all categories, the average accuracy of category  $A$  is the sum of the accuracies of all MRI images on the test set for category  $A$  over the number of all images containing

true annotations for category A. Following is the equation.

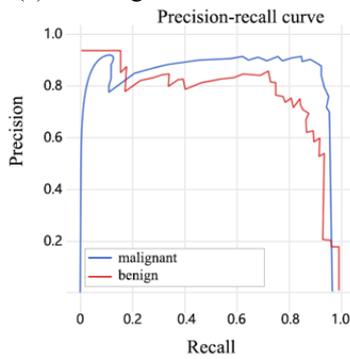
$$average\ precision_A = \frac{\sum precision_A}{N(total\ images)_A}$$

The mean average precision is then the expectation of the AP value for all categories, expressed by the following formula:

$$mean\ average\ precision = \frac{\sum average\ precision}{N(classes)}$$



(a) The original Faster R-CNN



(b) The improved model in this paper

**Figure 7.** PR curves for evaluating the performance of Faster R-CNN on MRI liver datasets

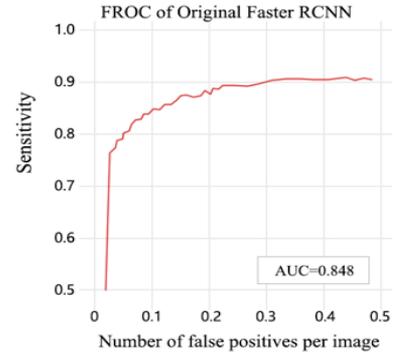
In addition, the magnitude of the mAP value is often calculated by plotting the precision-recall (PR) curve during the actual calculation. Figure 7 and Table 3 show the comparison of the evaluation results between the improved Faster R-CNN model and the original Faster R-CNN model in this paper on the constructed MRI liver occupancy dataset.

As shown in (a) of Figure 7, the original Faster R-CNN did not detect and classify benign occlusions well (AP=0.648) because the benign occupancies themselves had a small dataset and were not easily identified due to their small size. The original model, however, had a high detection accuracy for the majority of malignant tumours (AP=0.842), suggesting that the original Faster R-CNN was more impacted by inter-class imbalance. The improved model ((b) in Figure 7) used data augmentation to improve the interclass imbalance, used CondenseNet to improve the feature extraction performance, custom designed anchors to match the lesion size, and did transfer learning pre-training. After these, the improved model achieved a more balanced detection and classification performance for benign and malignant tumours, with the mAP value improving from the original 0.745 to 0.848 (see Table 3).

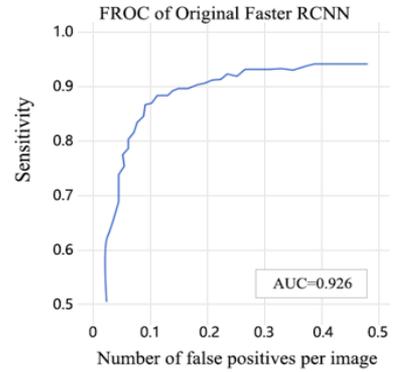
**Table 3.** Comparison of the mAP of the original Faster RCNN model and the improved model in this paper

Models	Benign occupancy (AP)	Malignant occupancy (AP)	mAP
Original Faster R-CNN model	0.648	0.842	0.745
Improved Faster R-CNN model	0.823	0.873	0.848

## 5.2 Receiver operating characteristic (ROC) curve



(a) FROC curve of the original Faster R-CNN model



(b) FROC curve of the improved Faster R-CNN model

**Figure 8.** FROC curves of the original Faster R-CNN model and the improved model in this paper

The ROC Area Under the Curve (AUC) is an evaluation metric frequently used in target detection classification. Due to the specificity of medical tasks, it is often necessary to obtain a high recall and sensitivity in prediction to avoid missing malignant patients, hence false positive predictors can be tolerated to some extent. Therefore, for the target detection problem on medical images, FROC, a variant of the ROC curve [34], is commonly used to evaluate the predictive performance of the model. FROC replaces the false positive rate on the horizontal axis with the mean number of false positives in the image, allowing the FROC curve to represent the level of recall and sensitivity that can be obtained at what level of false positives. Figure 8 shows the FROC curves and their AUC values for the original Faster R-CNN model and the improved model in this paper.

The FROC curves in Figure 8 above and the results in Table 4 are the prediction results based on a uniform distribution of 100 threshold points between 0 and 1 as IoU thresholds on the dataset of this paper. Compared with the original Faster R-CNN model (Sen = 0.912 when FP = 0.432), the improved model can obtain a higher sensitivity peak at a lower false positive rate (Sen = 0.948 when FP = 0.402), and the sensitivity of the improved model is higher than that of the

original Faster R-CNN model at the same false positive level. The sensitivity of the improved model was higher than that of the original Faster RCNN model at the same level of false positives. In addition, we extended the maximum value of the horizontal coordinate of the above FROC curve to 1 and kept the maximum value of the vertical coordinate unchanged to calculate the AUC value of the FROC curve. The corresponding AUC values of the original Faster R-CNN model and the improved model in this paper were 0.848 and 0.926, respectively. The above findings suggest that the improved Faster R-CNN model can help improve the performance of the Faster R-CNN model for liver occupancy detection and benign and malignant classification on MRI liver images.

**Table 4.** Comparison of the sensitivity of the original Faster R-CNN model and the improved model

	Original Faster R-CNN model	Improved Faster R-CNN model
Sensitivity of liver occupancy detection (FP=0.125)	0.846	0.891
Sensitivity of liver occupancy detection (FP=0.25)	0.879	0.928
Highest sensitivity for liver occupancy detection	0.912 (FP=0.432)	0.948 (FP=0.402)

## 6. SUMMARY

With the development of medical big data and MRI imaging technology, liver MRI has a higher sensitivity and cancer detection rate than CT examination, making early detection and diagnosis of liver cancer possible. Based on this, this paper first constructs a standard dataset for early detection and diagnosis of liver cancer in collaboration with relevant hospitals to overcome the current lack of MRI liver datasets in the field for carrying out research work on a computer-aided detection and diagnosis system for MRI of liver cancer. Wavelet-based soft threshold denoising is then used in the image pre-processing work to remove imaging thermal and physiological noise from the MRI images. The dataset is then annotated with the location of the lesion and its benignity or malignancy on each image, under the guidance of a specialist radiologist. In addition, to increase the data volume of the dataset, this paper uses an image geometric transformation to augment the original data, increasing the image texture information embedded in each image and the overall dataset data volume. The paper then proposes a computer-aided detection and diagnosis system based on the improved Faster R-CNN algorithm. The experimental comparison with the detection results of the original Faster R-CNN model demonstrates that the method in this paper achieves higher detection sensitivity on the constructed MRI standard dataset. This research paper provides a second suggestion to improve the efficiency of radiologists, meet the radiologists' need for reading images, and helps physicians for early diagnosis of liver cancer.

## ACKNOWLEDGMENT

This paper was supported by the Construction Fund of Key

medical disciplines of Hangzhou (Grant No.: OO20200265).

## REFERENCES

- [1] Wong, R.J., Ahmed, A. (2020). Understanding gaps in the hepatocellular carcinoma cascade of care: opportunities to improve hepatocellular carcinoma outcomes. *Journal of Clinical Gastroenterology*, 54(10): 850-856. <https://doi.org/10.1097/MCG.0000000000001422>
- [2] Heimbach, J.K., Kulik, L.M., Finn, R.S., Sirlin, C.B., Abecassis, M.M., Roberts, L.R., Marrero, J.A. (2018). AASLD guidelines for the treatment of hepatocellular carcinoma. *Hepatology*, 67(1): 358-380. <https://doi.org/10.1002/hep.29086>
- [3] Zhang, C.H., Ni, X.C., Chen, B.Y., Qiu, S.J., Zhu, Y.M., Luo, M. (2019). Combined preoperative albumin-bilirubin (ALBI) and serum  $\gamma$ -glutamyl transpeptidase (GGT) predicts the outcome of hepatocellular carcinoma patients following hepatic resection. *Journal of Cancer*, 10(20): 4836-4845. <https://doi.org/10.7150/jca.33877>
- [4] Medical Administration and Hospital Administration of the National Health Commission of the People's Republic of China. (2019). Guidelines for the diagnosis and treatment of primary liver cancer. *Chinese Journal of Liver Diseases*, 28(2): 112-128.
- [5] Wang, H., Naghavi, M., Allen, C., Barber, R.M., Bhutta, Z.A., Carter, A., Bell, M.L. (2016). Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980-2015: A systematic analysis for the Global Burden of Disease Study 2015. *The Lancet*, 388(10053): 1459-1544. [https://doi.org/10.1016/S0140-6736\(16\)31012-1](https://doi.org/10.1016/S0140-6736(16)31012-1)
- [6] Chen, W., Zheng, R., Baade, P.D., Zhang, S., Zeng, H., Bray, F., He, J. (2016). Cancer statistics in China, 2015. *CA: A Cancer Journal for Clinicians*, 66(2): 115-132. <https://doi.org/10.3322/caac.21338>
- [7] Elsayes, K.M., Hooker, J.C., Agrons, M.M., Kielar, A.Z., Tang, A., Fowler, K.J., Sirlin, C.B. (2017). 2017 Version of LI-RADS for CT and MR Imaging: An Update. *Radiographics: A Review Publication of the Radiological Society of North America, Inc*, 37(7): 1994-2017. <https://doi.org/10.1148/rq.2017170098>
- [8] Ayuso, C., Rimola, J., Vilana, R., Burrel, M., Darnell, A., García-Criado, Á., Brú, C. (2018). Diagnosis and staging of hepatocellular carcinoma (HCC): Current guidelines. *European Journal of Radiology*, 101: 72-81. <https://doi.org/10.1016/j.ejrad.2018.01.025>
- [9] Marrero, J.A., Kulik, L.M., Sirlin, C.B., Zhu, A.X., Finn, R.S., Abecassis, M.M., Heimbach, J.K. (2018). Diagnosis, staging, and management of hepatocellular carcinoma: 2018 practice guidance by the American association for the study of liver diseases. *Hepatology*, 68(2): 723-750. <https://doi.org/10.1002/hep.29913>
- [10] Choi, J.Y., Cho, H.C., Sun, M., Kim, H.C., Sirlin, C.B. (2013). Indeterminate observations (liver imaging reporting and data system category 3) on MRI in the cirrhotic liver: fate and clinical implications. *American Journal of Roentgenology*, 201(5): 993-1001. <https://doi.org/10.2214/ajr.12.10007>
- [11] Jiang, J., Wang, W., Cui, Y.N., Zhang, M.W., Chen, D., Fang, X., Liu, A.L. (2021). Evaluation of the diagnostic value of CT and MRI for hepatocellular carcinoma less

- than or equal to 3 cm based on the 2018 version of the liver imaging report and data system. *Magnetic Resonance Imaging*, 12(9): 25-29, 44. <https://doi.org/10.12015/issn.1674-8034.2021.09.006>
- [12] Meng, L., Tian, Y., Bu, S. (2020). Liver tumor segmentation based on 3D convolutional neural network with dual scale. *Journal of Applied Clinical Medical Physics*, 21(1): 144-157. <https://doi.org/10.1002/acm2.12784>
- [13] Tang, W., Zou, D., Yang, S., Shi, J., Dan, J., Song, G. (2020). A two-stage approach for automatic liver segmentation with Faster R-CNN and DeepLab. *Neural Computing and Applications*, 32(11): 6769-6778. <https://doi.org/10.1007/s00521-019-04700-0>
- [14] Li, X., Chen, H., Qi, X., Dou, Q., Fu, C.W., Heng, P.A. (2018). H-DenseUNet: Hybrid densely connected UNet for liver and tumor segmentation from CT volumes. *IEEE Transactions on Medical Imaging*, 37(12): 2663-2674. <https://doi.org/10.1109/TMI.2018.2845918>
- [15] Bousabarah, K., Letzen, B., Tefera, J., Savic, L., Schobert, I., Schlachter, T., Lin, M. (2021). Automated detection and delineation of hepatocellular carcinoma on multiphasic contrast-enhanced MRI using deep learning. *Abdominal Radiology*, 46(1): 216-225. <https://doi.org/10.1007/s00261-020-02604-5>
- [16] Kim, J., Min, J.H., Kim, S.K., Shin, S.Y., Lee, M.W. (2020). Detection of hepatocellular carcinoma in contrast-enhanced magnetic resonance imaging using deep learning classifier: A multi-center retrospective study. *Scientific Reports*, 10(1): 1-11. <https://doi.org/10.1038/s41598-020-65875-4>
- [17] Zhao, J., Li, D., Kassam, Z., Howey, J., Chong, J., Chen, B., Li, S. (2020). Tripartite-GAN: Synthesizing liver contrast-enhanced MRI to improve tumor detection. *Medical Image Analysis*, 63: 101667. <https://doi.org/10.1016/j.media.2020.101667>
- [18] Li, C., Zhou, Y., Li, Y., Yang, S. (2021). A coarse-to-fine registration method for three-dimensional MR images. *Medical & Biological Engineering & Computing*, 59(2): 457-469. <https://doi.org/10.1007/s11517-021-02317-x>
- [19] Yang, Z.H., Feng, F., Wang, X.Y. (2010). Guidelines for magnetic resonance imaging techniques: Examination norms, clinical strategies, and new technologies (Revised Edition). *Chinese Journal of Medical Imaging*, 2010(4): 312.
- [20] Oladiran, O., Gichoya, J., Purkayastha, S. (2017). Conversion of JPG image into DICOM image format with one click tagging. In *International Conference on Digital Human Modeling and Applications in Health, Safety, Ergonomics and Risk Management*, pp. 61-70. [https://doi.org/10.1007/978-3-319-58466-9\\_6](https://doi.org/10.1007/978-3-319-58466-9_6)
- [21] Hadiyoso, S., Zakaria, H., Ong, P.A., Mengko, T.L.E.R. (2021). Hemispheric coherence analysis of wide band EEG signals for characterization of post-stroke patients with dementia. *Traitement du Signal*, 38(4): 985-992. <https://doi.org/10.18280/ts.380408>
- [22] Das, A., Agrawal, S., Samantaray, L., Panda, R., Abraham, A. (2020). State-of-the art optimal multilevel thresholding methods for brain MR image analysis. *Revue d'Intelligence Artificielle*, 34(3): 243-256. <https://doi.org/10.18280/ria.340302>
- [23] Pal, C., Das, P., Chakrabarti, A., Ghosh, R. (2017). Rician noise removal in magnitude MRI images using efficient anisotropic diffusion filtering. *International Journal of Imaging Systems and Technology*, 27(3): 248-264. <https://doi.org/10.1002/ima.22230>
- [24] Ismael, A.A., Baykara, M. (2021). Digital image denoising techniques based on multi-resolution wavelet domain with spatial filters: A review. *Traitement du Signal*, 38(3): 639-651. <https://doi.org/10.18280/ts.380311>
- [25] Zhou, Y., Liu, W.P., Luo, Y.Q., Zong, S.X. (2021). Small object detection for infected trees based on the deep learning method. *Scientia Silvae Sinicae*, 57(3): 98-107.
- [26] Ge, C., Gu, I.Y.H., Jakola, A.S., Yang, J. (2020). Enlarged training dataset by pairwise GANs for molecular-based brain tumor classification. *IEEE Access*, 8: 22560-22570. <https://doi.org/10.1109/ACCESS.2020.2969805>
- [27] Yang, X., Liu, L., Li, T. (2022). MR-UNet: An UNet model using multi-scale and residual convolutions for retinal vessel segmentation. *International Journal of Imaging Systems and Technology*, 32(5): 1588-1603. <https://doi.org/10.1002/ima.22728>
- [28] Cai, S., Wu, Y., Chen, G. (2022). A novel elastomeric UNet for medical image segmentation. *Frontiers in Aging Neuroscience*, 14: 841297. <https://doi.org/10.3389/fnagi.2022.841297>
- [29] Ren, S., He, K., Girshick, R., Sun, J. (2015). Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6): 1137-1149. <https://doi.org/10.1109/TPAMI.2016.2577031>
- [30] Uemura, T., Näppi, J.J., Hironaka, T., Kim, H., Yoshida, H. (2020). Comparative performance of 3D-DenseNet, 3D-ResNet, and 3D-VGG models in polyp detection for CT colonography. In *Medical Imaging 2020: Computer-Aided Diagnosis*, 11314: 736-741. <https://doi.org/10.1117/12.2549103>
- [31] Huang, G., Liu, S., Van der Maaten, L., Weinberger, K.Q. (2018). Condensenet: An efficient densenet using learned group convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2752-2761. <https://doi.org/10.1109/CVPR.2018.00291>
- [32] Hasan, S.K., Linte, C.A. (2020). CondenseUNet: a memory-efficient condensely-connected architecture for bi-ventricular blood pool and myocardium segmentation. In *Medical Imaging 2020: Image-Guided Procedures, Robotic Interventions, and Modeling*, 11315: 402-408. <https://doi.org/10.1117/12.2550640>
- [33] Everingham, M., Van Gool, L., Williams, C.K., Winn, J., Zisserman, A. (2010). The pascal visual object classes (VOC) challenge. *International Journal of Computer Vision*, 88(2): 303-338. <https://doi.org/10.1007/s11263-009-0275-4>
- [34] Bandos, A.I., Obuchowski, N.A. (2019). Evaluation of diagnostic accuracy in free-response detection-localization tasks using ROC tools. *Statistical Methods in Medical Research*, 28(6): 1808-1825. <https://doi.org/10.1177/0962280218776683>