Original Article

Automated liver tumor detection in abdominal ultrasonography with a modified faster region-based convolutional neural networks (Faster R-CNN) architecture

Kenji Karako[1,2#], Yuichiro Mihara[2#], Junichi Arita[2], Akihiko Ichida[2], Sung Kwan Bae[2], Yoshikuni Kawaguchi[2], Takeaki Ishizawa[2], Nobuhisa Akamatsu[2], Junichi Kaneko[2], Kiyoshi Hasegawa[2], Yu Chen[1]

[1]Department of Human and Engineered Environmental Studies, Graduate School of Frontier Sciences, The University of Tokyo, Chiba, Japan; [2]Artificial Organ and Transplantation Surgery Division, Department of Surgery, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan

Contributions: (I) Conception and design: K Karako, Y Mihara, K Hasegawa, Y Chen; (II) Administrative support: J Arita, K Hasegawa, Y Chen; (III) Provision of study materials or patients: Y Mihara, J Arita, A Ichida, SK Bae, Y Kawaguchi, T Ishizawa, N Akamatsu, J Kaneko; (IV) Collection and assembly of data: Y Mihara, J Arita, A Ichida, SK Bae, Y Kawaguchi, T Ishizawa, N Akamatsu, J Kaneko; (V) Data analysis and interpretation: K Karako, Y Mihara; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

[#]These authors contributed equally to this work.

Correspondence to: Yu Chen. Department of Human and Engineered Environmental Studies, Graduate School of Frontier Sciences, The University of Tokyo, 5-1-5 Kashiwa-no-ha, Kashiwa, Chiba 227-8568, Japan. Email: chen@edu.k.u-tokyo.ac.jp; Kiyoshi Hasegawa. Artificial Organ and Transplantation Surgery Division, Department of Surgery, Graduate School of Medicine, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8655, Japan. Email: kihase-tky@umin.ac.jp.

Background: Although diagnostic ultrasound can non-invasively capture the image of abdominal viscera, diagnosis of the continuous ultrasound liver images to detect a liver tumor effectively and to determine whether the detected is benign or malignant is nontrivial. In order to minimize the gaps in diagnostic accuracy depending on doctor's proficiency, we built an automated system to support the ultrasonography of liver tumors by employing deep learning technologies.

Methods: We constructed a neural network model for the automated detection of tumor tissues and blood vessels from the sequential liver ultrasound images. Faster region-based convolutional neural networks (Faster R-CNN) is employed as a base model for the object detection, which can output the detection results in 4 frames per second and enable the system to be particularly suitable for the real time ultrasonography. Moreover, we proposed a new neural network architecture feeding both the current and previous images into Faster R-CNN. For training the models, intraoperative ultrasound images obtained from one hepatocellular carcinoma (HCC) patient were used. The obtained image was a multifaceted observation of the liver and includes one HCC and some blood vessels. We labeled 91 images with the help of a liver specialist. We compared the tumor detection performance of the plain Faster R-CNN model with that of the proposed model.

Results: We find that both the models performed well in detecting HCC and blood vessels, after training with 400 epochs using Adam. However, the mean precision of our model reaches 0.549, which is 0.019 better than that of the plain Faster R-CNN, and the mean sensitivity of our model about HCC reaches 0.623±0.385 for 30 scenes of sequential liver ultrasound images, which is also 0.146 better than that of the plain Faster R-CNN model.

Conclusions: The comparison between the proposed model and the plain Faster R-CNN model shows that we achieved better accuracy in tumor detection, in terms of the mean precision as well as the mean sensitivity, with the proposed model.

Keywords: Liver; deep learning; ultrasonography; object detection; hepatocellular carcinoma (HCC)

Submitted Jan 28, 2021. Accepted for publication Apr 14, 2021.
doi: 10.21037/hbsn-21-43
View this article at: http://dx.doi.org/10.21037/hbsn-21-43

Introduction

In an era where deep learning is widely applied to the image recognition (1-3), the technology has attracted attention in the medical field as well. The use of deep learning models for the automated diagnosis based on computed tomography (CT) (4-6) or magnetic resonance imaging (MRI) (7-9) images has been proposed in several literatures. On the other hands, there are some studies of the classification based on ultrasound images of thyroid nodules (10-12), breast cancer (13,14), and liver fibrosis (15). The reason why deep learning is difficult to apply to ultrasonography mainly lies in that the dynamic images obtained from ultrasonography are a series of cross-sectional objects, whereas most research on deep learning image recognition focused on images acquired optically outside an object. Dynamic cross-sectional images taken during ultrasonography differ very much depending on the angle and the position of ultrasound probe, a fact which requires improvements of the existing deep learning technology in order to extract features from images with diverse qualities and characteristics. In addition, unlike in a CT or MRI scan, position and angle of the ultrasound probe also strongly depend on the operation method of the doctor and the state of the patient. Therefore, more versatile training data and novel neural network architectures are needed for the application of deep learning in ultrasonography than in CT or MRI.

The current study concerns the automated diagnosis of a liver tumor based on ultrasonography. As hepatocellular carcinoma (HCC) is a major tumor of the liver with no early symptoms, a regular examination for HCC is recommended. Being a non-invasive and simple method, ultrasonography is a more popular examination for HCC as compared with CT or MRI. On the other hand, identifying HCC based on ultrasonography requires expertise and experience, and its accuracy also depends on the skill and experience of the doctor to a large extent. In other words, only with the treatment of a skilled and experienced doctor can the accuracy of diagnosis be ensured in the ultrasonography examination. The instability of ultrasonography takes root in the fact that a doctor needs to visualize internally the three-dimensional liver through the ultrasound monochrome images, which are dynamically changing cross-sectional images of the liver. Recognizing an object from its cross-sectional liver images means properly mapping those images into tumor tissues, blood vessels, and other areas of the liver with rich physiological knowledge.

The aim of our study is to build a deep learning system that can detect the liver tumor from ultrasound images. We present the following article in accordance with the STARD reporting checklist (available at https://hbsn.amegroups.com/article/view/10.21037/hbsn-21-43/rc).

Methods

Section "A model of tumor detection based on two sequential images" briefly describes the tumor detection model. Faster R-CNN (16) and its three component layers are summarized in section "Basic object detection framework: Faster R-CNN". The aggregation framework based on two sequential images is described in section "Aggregation framework based on two sequential frames". The dataset for tumor detection was described in section "Dataset". Training settings and the evaluation method are described in section "Training settings and evaluation method".

A model of tumor detection based on two sequential images

Our purpose is to highlight the location of tumors and blood vessels with differently colored square frames on ultrasound video in real time. To achieve this aim, we created a model of tumor detection based on two sequential images. The base framework for our model is Faster R-CNN (16), a deep learning object detection model that can process 4 frames per second. Extra layers are added and information is fed into Faster R-CNN in order to catch three-dimensional features of objects from two sequential cross-section images. *Figure 1* shows an overview of the framework. When recognizing a tumor from ultrasound video, essentially the doctor reconstructs the three-dimensional structure of the tumor in his or her own mind from continuous changes in the video. Expecting that deep learning mimics the above recognition process, we constructed a model capturing sequential images and the change in the time direction. Given two sequential images, aggregation layers will extract the three-dimensional

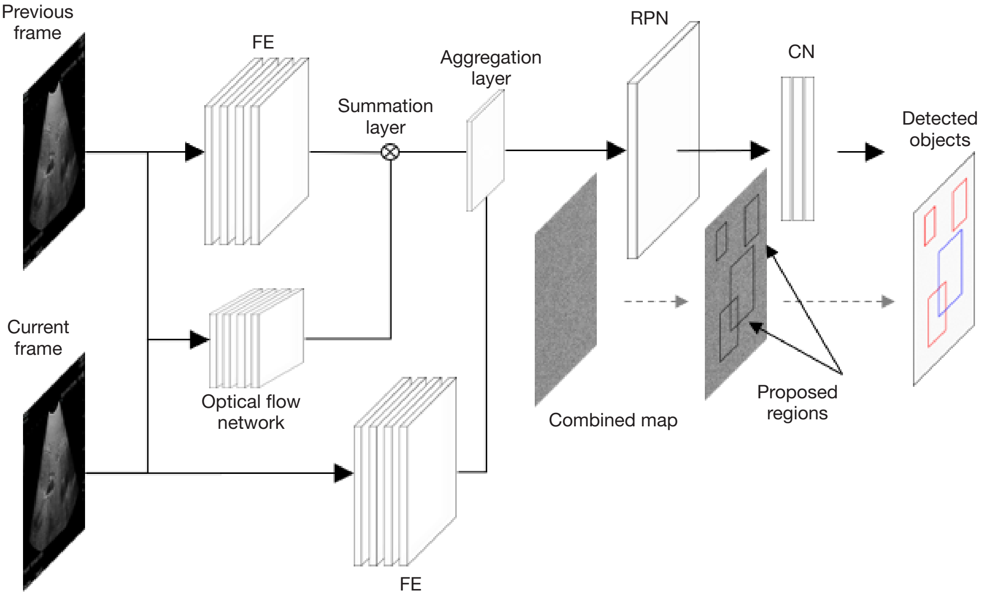

Figure 1 Overview of a model of tumor detection based on two sequential images. FE, Feature Extractor; RPN, Region Proposal Network; CN, Classification Network.

features by combining features in the previous image, the current image, and the optical flow information which is the change in the time direction estimated from the previous and current images. The Region Proposal Network (RPN) and Classification Network (CN) are responsible for predicting the precise position as well as labeling objects based on the three-dimensional features obtained earlier.

Basic object detection framework: Faster R-CNN

Faster R-CNN is a well-known method of object detection. Shown in *Figure 2*, it consists of three components: a Feature Extractor (FE), a RPN, and a CN. All the three component layers have a neural network from end to end, enabling faster processing of information. The role of each layer is described below.

FE

FE consists of a deeply connected convolutional neural network. Given an image input, FE outputs a map of feature vectors expressing objects in the image. The feature vector is represented by a multidimensional vector based on information near each pixel of the object. In order to extract the feature vectors, FE must be trained with a large number of images. As there are not enough ultrasound images to train FE in this study, the pretrained network "resnet50" (17) which was trained with the ImageNet classification dataset (18) is used as the initial setting.

RPN and CN

RPN is a convolutional neural network which identifies regions containing the objects in the image from a feature vector map output from FE. On the other hand, CN is a fully connected neural network that predicts the label of an object as well as the precise area where objects may be found based on proposed regions and the feature vector map. Like FE, RPN and CN need to be trained with a large amount of data to perform their tasks, hence the pretrained networks are used as initial settings.

Aggregation framework based on two sequential frames

We added new layers to obtain three-dimensional features by synthesizing two-dimensional features extracted from the sequential ultrasound images. As shown in *Figure 1*, the aggregation framework consists of five parts: a previous FE, a current FE, an optical flow network, a summation layer, and an aggregation layer. The previous and current FEs use the same networks as the FE described above and extract features from the previous and current frame. The optical flow network is an end-to-end convolutional neural network that maps the two-dimensional movement vector for each pixel in the previous and current frame. The optical flow network has been proposed as FlowNet in (19) and several improved models have been proposed (20,21). In this study, we used pretrained LiteFlowNet (22), which can quickly map movement vectors than other FlowNets. *Figure 3* shows a sample of obtained optical flow map from the sequential ultrasound images. The map was generated as an optical movement vector of each pixel in the two-dimensional direction. The optical flow map in *Figure 3* is an image in which the movement amount vector is converted into an RGB color model with the amount of change in the horizontal direction as red, the amount of change in the vertical direction as green, and blue as a fixed value of 111. The red and green code values indicate a positive direction if they are larger than 111. It can be seen that the yellow blood vessels in the center of the optical flow map are moving in the lower right direction. The summation layer and aggregation layer are networks to synthesize the features map obtained from the aforementioned networks and to extract a three-dimensional features map. Synthesis consists of two steps. First, the summation layer adds a movement vector map to the previous features map as different channels for each coordinate. Second, the aggregation layer uses a bottleneck structure (17,23) to combine the current features map and the previous features map. Through this processing, the summation layer outputs information in the previous

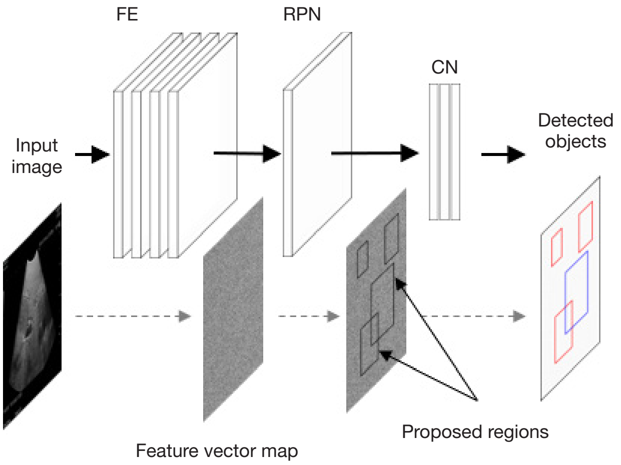

Figure 2 Overview of Faster R-CNN. FE, Feature Extractor; RPN, Region Proposal Network; CN, Classification Network; Faster R-CNN, faster region-based convolutional neural networks.

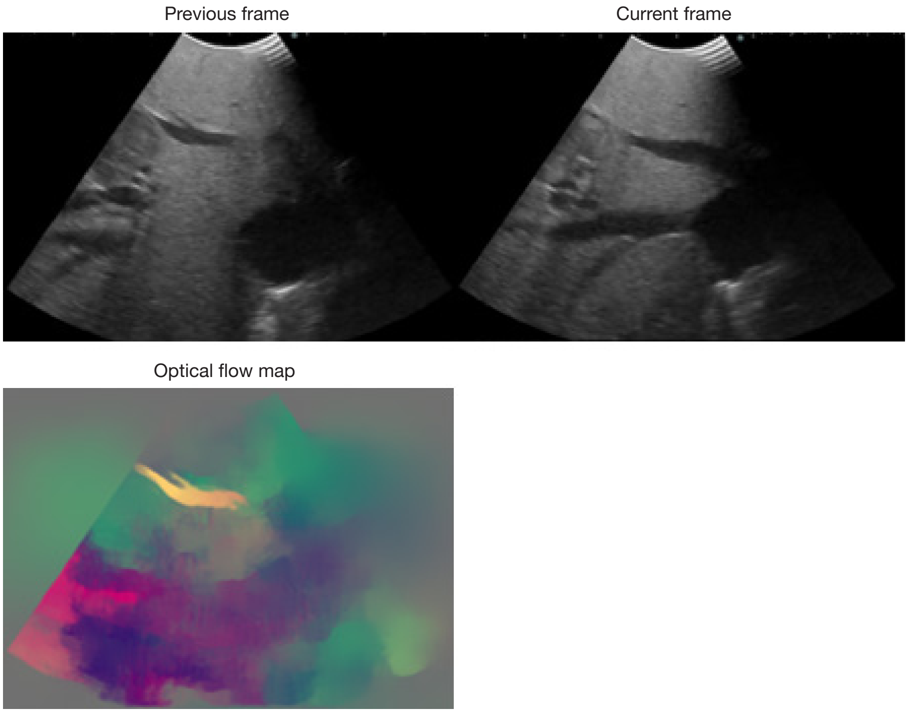

Figure 3 The sample of optical flow map obtained from two sequential images. The optical flow map was generated as an optical movement vector of each pixel in the two-dimensional direction from sequential images. Red color code indicates the amount of change in the horizontal direction. Green color code indicates the amount of change in the vertical direction. Blue color code is set as a fixed value of 111.

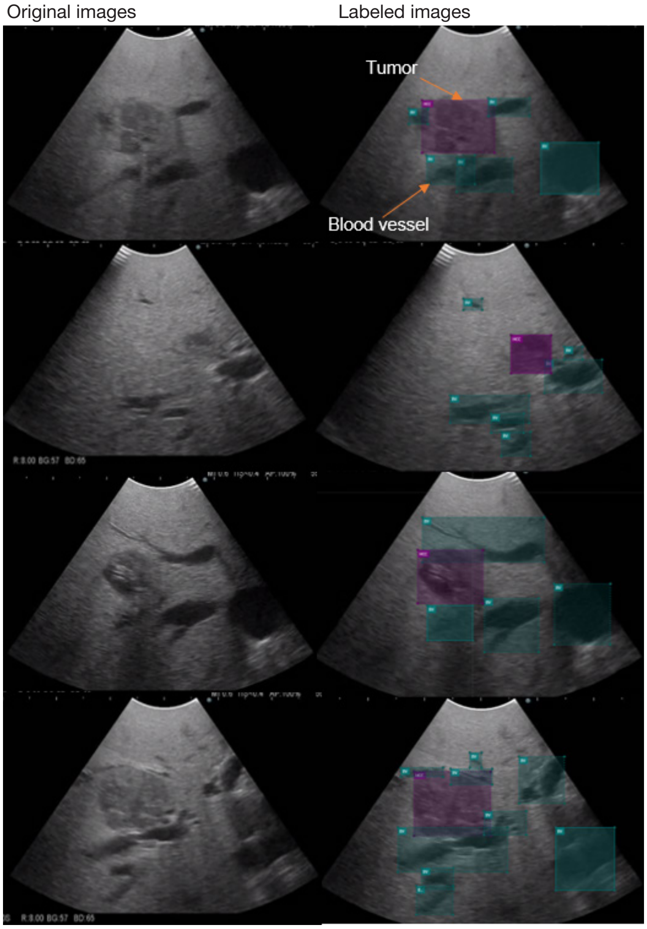

Figure 4 Examples of training images.

features map and movement information that is important to tumor detection. A features map including changes in three dimensions is generated by the aggregation layer and combined with the current features map.

Dataset

To evaluate our model, a training dataset is created from an ultrasound video of a patient with HCC at The University of Tokyo Hospital. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). This study was approved by the appropriate institutional review board of Graduate School of Medicine and Faculty of Medicine, The University of Tokyo (No. 2019166NI) and informed consent was obtained in the form of opt-out on the website. The video is a 2 min 58 s video taken during an intraoperative ultrasonography performed during surgery on one HCC patient. The video contained one tumor and its size was about 2 cm. We picked some scenes, which are 1.1 s in total, from the video where tumors and blood vessels were continuously visible and cropped. Ultrasound images from the videos were labeled with a rectangular frame and classified as either tumor or blood vessel, as shown in *Figure 4*. Labeling for tumor and anatomical structure was done with experienced HPB surgeon. The location of tumors was recognized by preoperative contrast-enhancement CT (CE-CT) images, EOB-MRI images, intraoperative ultrasound (IOUS) images and post-operative pathological findings. The whole labeled data were obtained from 91 images. Seventy-two images of the dataset were used to train the model and 19 images were used to evaluate the model.

Training settings and evaluation method

To realize the tumor detection, it is necessary to repeatedly train the model with the correct data. The procedures of training are as follows. The first step is inputting the image of the training dataset to the model and predicting the coordinates and labels of the objects in the image. The second step is calculating the error between the correct outputs and predicted outputs. The third step is optimization of the parameters to reduce the error. We used Adam (24), which is often used in deep learning, as the optimization method. Adam is a gradient descent algorithm with an adaptive momentum that computes adaptive learning rates for each parameter of neural networks. These processes are called epoch. The error usually becomes smaller through multiple epochs. However, this only indicates that the trained model can detect the objects with high accuracy for known data. Therefore, we train the model with the 72 images divided for training in section "Dataset" as the training dataset and evaluate COCO mean average precision (mAP) (25) for 19 images divided for evaluation in section "Dataset" as the heretofore unseen dataset. As for the evaluation of accuracy, mAP is a general unitless metric for the object detection considering not only the accuracy of the label of object predicted by the model but also the accuracy of the detected area. The mAP measures Average Precision for intersection over union (IoU) from 0.5 to 0.95 with a step size of 0.05, which is the actual metric for object detection. There are other metrics such as average precision at IoU =0.50 (AP_{50}) and average precision at IoU =0.75 (AP_{75}), which are less strict than mAP.

Table 1 Comparison of AP in tumor detection

Accuracy	Methods	
	Our model	Faster R-CNN
Mean AP	0.549	0.530
AP_{50}	0.887	0.881
AP_{75}	0.641	0.625

AP, average precision; Faster R-CNN, faster region-based convolutional neural networks.

Table 2 Comparison of mean sensitivity and detection results by paired two-samples *t*-test

Label	Mean sensitivity per scene		P value	*t*-value
	Our model	Faster R-CNN		
Tumor	0.627±0.390	0.456±0.411	0.003	–2.921
Blood vessel	0.558±0.201	0.576±0.202	0.290	–0.559

Faster R-CNN, faster region-based convolutional neural networks.

Results

Performance evaluation

We trained our model and Faster R-CNN with 400 epochs using Adam with learning rates starting at 1×10^{-3}. The evaluation results are shown in *Table 1*. The results are the best values in 400 epochs. Each model performed well at detecting tumors and blood vessels. The mAP of our model reaches 0.549, which is 0.019 better than the average accuracy of Faster R-CNN.

Comparison of sensitivity by label

Our model showed better results in mAP for evaluation data in section “Performance evaluation”. However, it is difficult to accurately evaluate the generalization performance of the models because the data collected in this study is obtained from only one patient and the size of data is small. For model improvement and performance comparison, we applied the two top-performing models obtained in section “Performance evaluation” to 30 scenes extracted from the video mentioned in section “Dataset”. Each scene has 5 images that were not included in the training and evaluation data. The total number of the images reaches 150. These images contain 142 tumors and 689 blood vessels. We applied each model to the 30 scenes and tallied the object count as 1 whenever IoU between true area and predicted area is 0.5 or more while the label is also correct. We calculated sensitivities for each label per scene. The mean sensitivity per scene of each model and the paired two-samples *t*-test are shown in *Table 2*. The mean sensitivity per scene of our model reaches 0.627±0.390, which is 0.171 better than that of Faster R-CNN. The P value in tumor detection performance is 0.003. Lastly, results of tumor detections predicted from some scenes are shown in Figure S1.

Discussion

The system built in this study focuses on morphological distinguishment between tumor tissues and blood vessels in the ultrasound images. Many object detection models, such as Faster R-CNN (16), Single Shot Detector (SSD) (26), and You Only Look Once (YOLO) (27), have been proposed for general object recognition and shown excellent results in detecting people or cars. The processing speed of Faster R-CNN is 4 frames per second, slower than the other methods though, can determine the position of the object with a higher accuracy. Furthermore, we can detect the specific area containing the object like segmentation by extending Faster R-CNN to Mask R-CNN (28). Hence, we used Faster R-CNN as the baseline framework in this study. The system will enable a simple and remote examination of HCC from ultrasound images, without consulting a liver specialist. Hence, it can be expected that the early detection of HCC may be realized with such a system in the near future.

As a related task, there are the detection and classification of thyroid nodules. In thyroid nodule recognition in ultrasound images, deep learning model based on YOLOv2 (29) showed performances comparable to experienced radiologists (10). Unlike the thyroid ultrasonography, liver ultrasonography is often observed from multiple angles, and the viewpoints changes significantly. Therefore, it would also be difficult to apply the method to ultrasound liver images in a straightforward manner, because object detection model such as YOLOv2 and Faster R-CNN used to be employed in the identification of two-dimensional features from a single image. To capture features of dynamic ultrasound images, a new object detection model has to be constructed for extracting the three-dimensional features from the dynamic ultrasound images.

Our model showed better performance than Faster R-CNN does in terms of mAP, which is a general criterion for object detection. A comparison of sensitivity to heretofore unseen sequential images in section "Comparison of sensitivity by label" shows that the performance of our model in the detection of tumor is higher. In the paired two-samples *t*-test, the P value in tumor detection performance was below 0.05, the standard of a statistical difference. On the other hand, there was no significant difference between both models in the sensitivity of blood vessels. Rather, our model showed a tendency to detect one blood vessel as a plurality of objects in Figure S1. The reason that our model showed the better performance only in detecting tumors is that the shape and position of the tumor do not change much between the two sequential frames, while the shape and position of the blood vessels vary significantly. When the change of the object between two frames is small, the aggregation framework can easily recognize the same object in the two frames as the same and make the features map reflect this recognition. On the other hand, if the change of the object in two frames is large, the same object may be recognized as different objects, hence the detection is not successful. This phenomenon can be seen from the result that one blood vessel is detected as a plurality of objects in other scenes of Figure S1. In addition, as a common trend in detections, shadows are mistakenly recognized as blood vessels in both models, and blood vessels outside the liver region are also detected.

In section "Results", 72 images were used as training data, and the total of 169 unlearned images were used as evaluation data. The evaluation data is used more than twice as much as the training data, and it is considered that the amount of data is sufficient for evaluating the detection performance of the ultrasound image. However, there is few learning data to evaluate the generalization performance about the tumor detection and blood vessels are not included in the data. The amount of data used in the study of Faster R-CNN and the studies applying image recognition to CT (4-6) or MRI (7-9) is enormous, exceeded at least triple-digits. As a study using deep learning related tumor detection with dynamic ultrasound images, there are erosions and ulcerations detection model trained with 440 dynamic images of wireless capsule endoscopy (30) and breast cancer detection model trained with 8,145 images of ultrasonography (31). Comparing these studies with our study, the amount of data is overwhelmingly insufficient. A large amount of diverse data is usually required for deep learning, and it is only possible to effectively recognize unlearned data by learning the features of general objects from the data. In addition, there is a lack of diversity because the data used in our study was obtained from one patient. Training the detection model with dynamic ultrasound images of the liver requires more diverse data than the CT and MRI. Usually, when performing an ultrasound examination of the liver, the viewpoint of ultrasonography differs depending on the operator and the position of the tumor. For more accurate tumor detection, it is desirable to collect not only images of liver tumors but also images of the liver and extrahepatic object taken in various situations from various angles. Moreover, the anatomical feature of liver and tumor appearance could be diverse by each patient. Thus, it is necessary to collect data from as many patients as possible to create the versatile system. Ensuring this diversity will be one of subject for future study. For additional learning data, it is appropriate to use images obtained from an intraoperative ultrasonography as we did in this study. Because an intraoperative ultrasonography is possible to obtain images from multiple directions without being disturbed by an intestinal tract, bones, and muscle tissues, compared with an extracorporeal ultrasonography. However, intraoperative ultrasonography, which can obtain a variety of liver images, may increase the cost of model training and training data creation. Meanwhile, there are data augmentation methods which can compensate for such shortage of data amount and diversity. We are planning to use Random Erasing (32) and noise addition which have been proposed as data augmentation for general still images. In addition, Generative Adversarial Network (GAN) (33) which generates similar data to learned data is also considered effective method for data augmentation. However, while GAN can generate synthetic dynamic ultrasound images, we basically require it to generate continuous ultrasound images as one scene. This is because it is not useful as learning data unless generating data similar to actual situation. There is the study of GAN (34) for the generation of continuous data. In that study, by inputting the facial expression into a model which has been trained with images of a particular person, images of that person having the facial expression can be created. By applying this method into the ultrasound image of the liver, continuous ultrasound may be created by inputting the tumor position which is designated continuously. If it can be applied to ultrasound images, we shall solve the problem of lack of data.

In conclusion, we developed a deep learning model using to detect a liver tumor in dynamic ultrasound images. In situations involving a limited amount of data, our model performs better than Faster R-CNN. In the future, we will collect more data on various liver tumors and improve our model to detect tumors and detailed areas.

Acknowledgments

Funding: This work was supported by JSPS KAKENHI (20K20214) to YM; and a grants-in-aid of the 106th annual congress of JSS Memorial Surgical Research Fund, Tokyo, Japan to YM.

Footnote

Reporting Checklist: The authors have completed the STARD reporting checklist. Available at https://hbsn.amegroups.com/article/view/10.21037/hbsn-21-43/rc

Data Sharing Statement: Available at https://hbsn.amegroups.com/article/view/10.21037/hbsn-21-43/dss

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://hbsn.amegroups.com/article/view/10.21037/hbsn-21-43/coif). YM reports grants form JSPS KAKENHI and a grants-in-aid of the 106th annual congress of JSS Memorial Surgical Research Fund. KH serves as an unpaid editorial board member of *Hepatobiliary Surgery and Nutrition*. The other authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The study was approved by the appropriate institutional review board of Graduate School of Medicine and Faculty of Medicine, The University of Tokyo (No. 2019166NI) and informed consent was taken from all individual participants.

References

1. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Proceedings of the 25th International Conference on Neural Information Processing Systems. Lake Tahoe, NV, Dec. 2012:1097-105.
2. He K, Zhang X, Ren S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE Trans Pattern Anal Mach Intell 2015;37:1904-16.
3. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. Medical Image Computing and Computer-Assisted Intervention (MICCAI), Springer, LNCS, 2015;9351:234-41.
4. Sharma B, Venugopalan K. Classification of hematomas in brain CT images using neural network. 2014 International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT). 7-8 Feb. 2014; Ghaziabad, India. IEEE, 2014:41-6.
5. Teramoto A, Fujita H, Yamamuro O, et al. Automated detection of pulmonary nodules in PET/CT images: Ensemble false-positive reduction using a convolutional neural network technique. Med Phys 2016;43:2821-7.
6. Huang X, Shan J, Vaidya V. Lung nodule detection in CT using 3D convolutional neural networks. 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017). 18-21 April 2017; Melbourne, VIC, Australia. IEEE, 2017:379-383.
7. Shi J, Zheng X, Li Y, et al. Multimodal Neuroimaging Feature Learning With Multimodal Stacked Deep Polynomial Networks for Diagnosis of Alzheimer's Disease. IEEE J Biomed Health Inform 2018;22:173-83.
8. Pereira S, Pinto A, Alves V, et al. Brain Tumor Segmentation Using Convolutional Neural Networks in MRI Images. IEEE Trans Med Imaging 2016;35:1240-51.
9. Saman Sarraf, Ghassem Tofighi, for the Alzheimer's Disease Neuroimaging Initiative. DeepAD: Alzheimer's Disease Classification via Deep Convolutional Neural Networks using MRI and fMRI. bioRxiv 070441. doi: https://doi.org/10.1101/070441.
10. Wang L, Yang S, Yang S, et al. Automatic thyroid nodule recognition and diagnosis in ultrasound imaging with the YOLOv2 neural network. World J Surg Oncol 2019;17:12.
11. Chi J, Walia E, Babyn P, et al. Thyroid Nodule Classification in Ultrasound Images by Fine-Tuning Deep Convolutional Neural Network. J Digit Imaging 2017;30:477-86.
12. Persichetti A, Di Stasio E, Coccaro C, et al. Inter- and

Intraobserver Agreement in the Assessment of Thyroid Nodule Ultrasound Features and Classification Systems: A Blinded Multicenter Study. Thyroid 2020;30:237-42.
13. Xu Y, Wang Y, Yuan J, et al. Medical breast ultrasound image segmentation by machine learning. Ultrasonics 2019;91:1-9.
14. Tanaka H, Chiu SW, Watanabe T, et al. Computer-aided diagnosis system for breast ultrasound images using deep learning. Phys Med Biol 2019;64:235013.
15. Lee JH, Joo I, Kang TW, et al. Deep learning with ultrasonography: automated classification of liver fibrosis using a deep convolutional neural network. Eur Radiol 2020;30:1264-73.
16. Ren S, He K, Girshick R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks. Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 1, 2015:91-9.
17. He K, Zhang X, Ren S, et al. Deep residual learning for image recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 27-30 June 2016; Las Vegas, NV, USA. IEEE, 201:770-778.
18. Russakovsky O, Deng J, Su H, et al. ImageNet Large Scale Visual Recognition Challenge. Int J Comput Vis. 2015;115:211-52.
19. Dosovitskiy A, Fischer P, Ilg E, et al. FlowNet: Learning Optical Flow with Convolutional Networks. 2015 IEEE International Conference on Computer Vision (ICCV). 7-13 Dec. 2015; Santiago, Chile. IEEE, 2015:2758-66.
20. Ranjan A, Black MJ. Optical Flow Estimation Using a Spatial Pyramid Network. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 21-26 July 2017; Honolulu, HI, USA. IEEE, 2017:2720-9.
21. Ilg E, Mayer N, Saikia T, et al. FlowNet 2.0: Evolution of Optical Flow Estimation with Deep Networks. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 21-26 July 2017; Honolulu, HI, USA. IEEE, 2017:1647-55.
22. Hui T, Tang X, Loy CC. LiteFlowNet: A lightweight convolutional neural network for optical flow estimation. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2018:8981-9.
23. Zagoruyko S, Komodakis N. Wide residual networks. arXiv preprint arXiv:1605.07146, 2016.
24. Kingma DP, Ba J. Adam: A method for stochastic optimization. CoRR 2014;abs/1412.6980. arXiv:1412.6980.
25. COCO: Common Objects in Context. Available online: http://cocodataset.org/#detection-eval (accessed March 24, 2020)
26. Liu W, Anguelov D, Erhan D, et al. SSD: Single shot multibox detector. European conference on computer vision. Springer, Cham, 2016:21-37.
27. Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection. Proceedings of the IEEE conference on computer vision and pattern recognition. 2016:779-88.
28. He K, Gkioxari G, Dollár P, et al. Mask R-CNN. Proceedings of the IEEE international conference on computer vision. 2017:2961-9.
29. Redmon J, Farhadi A. YOLO9000: Better, Faster, Stronger. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 21-26 July 2017; Honolulu, HI, USA. IEEE, 2017:6517-25.
30. Aoki T, Yamada A, Aoyama K, et al. Automatic detection of erosions and ulcerations in wireless capsule endoscopy images based on a deep convolutional neural network. Gastrointest Endosc 2019;89:357-63.e2.
31. Qi X, Zhang L, Chen Y, et al. Automated diagnosis of breast ultrasonography images using deep neural networks. Med Image Anal 2019;52:185-98.
32. Zhong Z, Zheng L, Kang G, et al. Random Erasing Data Augmentation. Proceedings of the AAAI Conference on Artificial Intelligence, 2020;34:13001-8.
33. Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets. 2014 Neural Information Processing Systems 2014;27:2672-80.
34. Otberdout N, Daoudi M, Kacem A, et al. Dynamic facial expression generation on Hilbert hypersphere with conditional Wasserstein generative adversarial nets. IEEE Trans Pattern Anal Mach Intell 2022;44:848-63

Cite this article as: Karako K, Mihara Y, Arita J, Ichida A, Bae SK, Kawaguchi Y, Ishizawa T, Akamatsu N, Kaneko J, Hasegawa K, Chen Y. Automated liver tumor detection in abdominal ultrasonography with a modified faster region-based convolutional neural networks (Faster R-CNN) architecture. HepatoBiliary Surg Nutr 2022;11(5):675-683. doi: 10.21037/hbsn-21-43

Supplementary

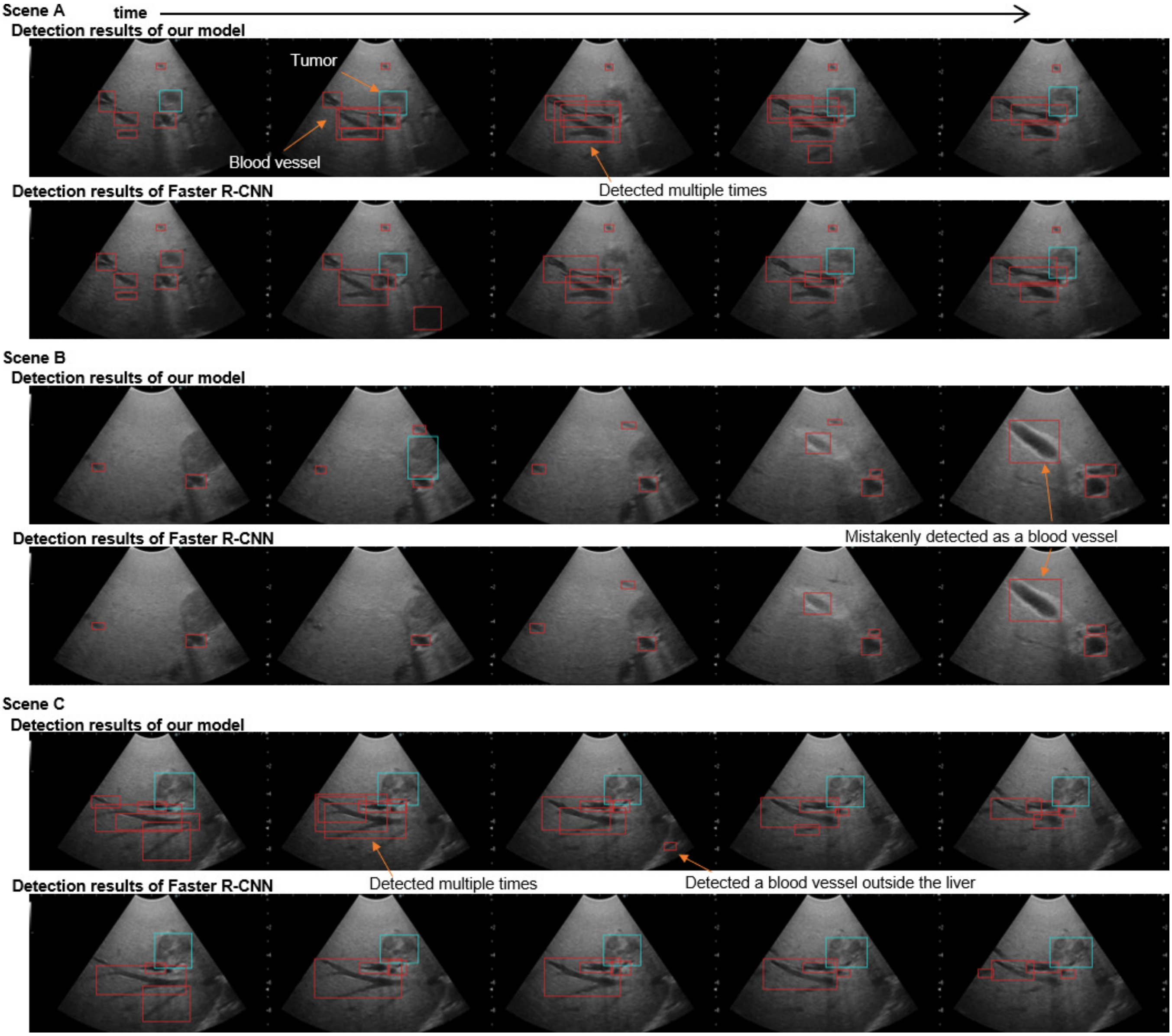
Scene A
time
Detection results of our model
Tumor
Blood vessel
Detection results of Faster R-CNN
Detected multiple times
Scene B
Detection results of our model
Detection results of Faster R-CNN
Mistakenly detected as a blood vessel
Scene C
Detection results of our model
Detection results of Faster R-CNN
Detected multiple times
Detected a blood vessel outside the liver

Figure S1 Tumor detection results of our model and Faster R-CNN for some heretofore unseen scenes. The box detected as an object by Classification Network with a probability of 50% or more were shown in the figure. The light blue box indicates the detected tumor. The red box indicates the detected blood vessel. Faster R-CNN, faster region-based convolutional neural networks.