Detection and Tracking of Liver Tumors for Ultrasound Diagnostic Support Using Deep Learning

Shoya Yamagishi, Keisuke Doman, and Yoshito Mekada
Graduate School of Engineering, Chukyo University, Toyota, Japan
Email: yamagishi.s2@md.sist.chukyo-u.ac.jp, {kdoman, y-mekada}@sist.chukyo-u.ac.jp

Naoshi Nishida and Masatoshi Kudo
Kindai University Faculty of Medicine, Osaka, Japan
Email: {naoshi, m-kudo}@med.kindai.ac.jp

***Abstract*—This paper proposes a diagnostic support method that simultaneously detects and tracks abdominal tumors from moving camera images taken by ultrasonography. Ultrasound diagnosis requires a doctor's experience because it is necessary to operate the equipment and perform the diagnosis simultaneously. Also, since the purpose of ultrasonic diagnosis is to find the presence of tumors, it is necessary to avoid overlooking them. We aim to develop a diagnostic support system that detects tumor cross-sections in each input frame using YOLOv3 and then tracks them using DeepSORT considering the instances of detected tumors. Here, even if a tumor disappears in an input frame due to doctor's probe manipulation, it can be identified as the same instance between the input frames and finally be grouped into the same cluster. If diagnostic support information can be ensembled for each cluster, it is expected that the accuracy of tumor detection should be improved. Experimental results showed that our system could perform the clustering of tumor cross-sections in real-time.**

***Index Terms*—liver tumor detection, tracking, ultrasound image, convolutional neural network**

I. Introduction

Diagnosis with an ultrasonic device is difficult because it requires a doctor to have much experience and also simultaneously perform probe manipulation and visual check. Therefore, it is desired to realize a computer-aided diagnosis system that can automatically detect tumors from ultrasonic images. We are studying the application of object detection algorithms to tumor detection from abdominal ultrasonography [1]. A naive way of feedback for detection results is to provide doctors with all the detected Regions of Interest (ROIs). There may be, however, a tumor instance detected multiple times through input frames, and also may be more than one ROI for one tumor image region. In such a case, just presenting all the detected ROIs is not a good way of feedback for detected results. This paper proposes a system that realizes abdominal tumor detection and its tracking using deep learning.

Various deep learning-based methods for object detection have been proposed. Some of them excel in real-time detection or specialize in detection accuracy [2]-[4]. There are studies to adjust the number of parameters based on their network models for each task [5] and also object tracking methods that can be used in conjunction with detection algorithms [6].

This paper includes three contributions: One is the development of a deep learning method for detecting tumor candidates from ultrasound images. Another is the development of a tracking process based on deep learning to group the candidate regions detected in each frame by the instance of tumor. The other is the development of a system that can provide grouped diagnostic information for each tumor instance by the detection and tracking methods.

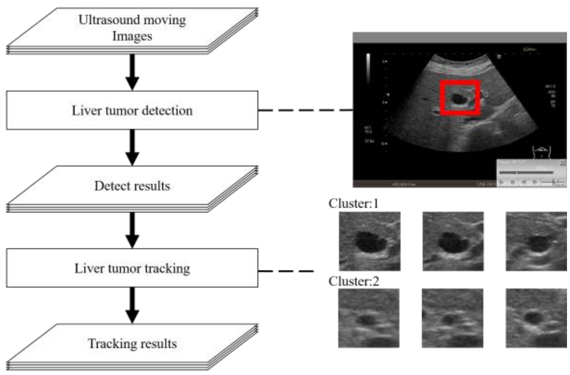


Figure 1. The liver tumor detection and tracking system process flow.

Fig. 1 shows the process flow of the system. The proposed system consists of two processes: 1) the detection of tumor regions and 2) the tracking of tumor regions as real time clustering of the detected regions.

II. Related Work

Object detection algorithms have been used in the medical field. Andréanne proposed a method to detect

Manuscript received July 21, 2021; revised December 12, 2021.

doi: 10.18178/joig.10.1.50-55

kidneys from CT images that show kidneys with a variety of forms, textures, positionings and contrasts [7]. Zhenxiao *et al.* improved YOLOv3 [2] to detect disc herniation in real time with higher accuracy by reducing the amount of model parameters [5]. These studies showed that the object detection algorithms were effective for gray-scale medical images such as CT and MRI images. Fig. 2 shows an example of Hepatocellular Carcinoma (HCC) with various features appearing in an ultrasound image. Since our dataset consists of tumors with various features, we expect that object detection algorithms based on deep learning are effective.

In our previous study by Nakashima *et al.* used an object detection algorithm to detect liver tumors [1]. In their experiments, they used YOLOv3 [2] or Faster R-CNN [4] for detection and ResNeXt50 [8] for false-positive elimination to improve the accuracy. As a result, the method, which combined object detection by YOLOv3 and false-positive elimination by ResNeXt50, had the highest accuracy and F-measure of 0.7891. Narinder *et al.* combined object detection algorithms with DeepSORT [6] to track people [9]. They investigated the detection performance in the case of individually using YOLOv3, Faster R-CNN, or SSD [3]. The mAP/framerate of each combination was as follows: 0.846/23 for YOLOv3, 0.969/3 for Faster R-CNN, and 0.691/10 for SSD. Faster R-CNN had the best accuracy but the lowest framerate. We concluded that YOLOv3, which achieved the second highest accuracy and the highest framerate, is reasonable for our system.

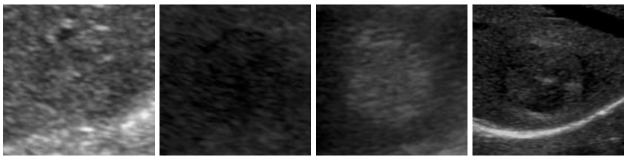


Figure 2. Examples of HCC several appearances in ultrasound images.

III. METHODS

Our method consists of two components: the detection of liver tumor regions using YOLOv3 and the tumor tracking using DeepSORT. The details of each component are described below.

A. Detection of Liver Tumor Regions

In the liver tumor detection process, we used a CNN-based object detection algorithm. Supposing that our system is simultaneously used with ultrasound examinations, more tumors may be acquired in real-time by using this system. One important thing here is to obtain candidate regions accurately in real time for several types of tumors with different appearances. As discussed in Section II, YOLOv3 is suitable for our system in terms of detection accuracy and computational efficiency.

B. Tumor Tracking

Our system uses DeepSORT for liver tumor tracking. The main motivation of introducing this process is to achieve grouping tumors whose cross-sectional shape changes by probe manipulation. Another motivation is to group tumors that temporarily disappear and reappear through input frames.

DeepSORT performs two types of grouping. The first is by using CNNs to obtain image features, which are then used for grouping during tracking. The second is by the SORT method [10], which is a component of DeepSORT. It uses only numerical information (the position and the size of bounding boxes) to predict the motion of objects between frames and group each bounding box, and can achieve high-speed processing.

IV. EXPERIMENTAL DATA

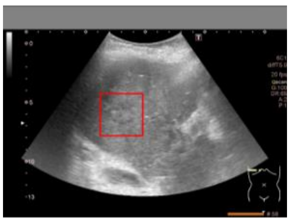
We used a dataset provided by the Japan Agency for Medical Research and Development (AMED). The dataset was constructed by collecting abdominal ultrasound images from multiple hospitals. All the cases were annotated with final diagnosis results and mainly consisted of HCC, metastatic, cyst, and hemangioma.

A. Image Dataset

Table I shows the details of the dataset. Fig. 3 shows an example of the dataset images. The dataset was composed of 74,460 images. Generally, liver tumors have several kinds of appearances in ultrasound images. Four-fifths, one-tenth, and the remaining of the images for each disease were used for training, validation, and test data of our system, respectively. Note that the images were extracted from video sequences, and thus, some of them had a near-duplicate appearance, for example, by selecting tumor images from adjacent video frames. We excluded near-duplicated images by evaluating the Humming distance of p-hash between the images.

TABLE I. THE DETAILS OF THE IMAGE DATASET

Diagnosis	# of images
Hepatic Cell Carcinoma	10,106
Metastatic Cancer	8,449
Hemangioma	23,993
Cyst	27,338
Other Tumor	4,574
Total	74,460



(a) Hepatic Cell Carcinoma (HCC)

(b) Metastatic Cancer

Figure 3. Examples of dataset used in the experiment.

B. Video Dataset

The details of the dataset are shown in Table II and Fig. 4 shows an example of a set of images in the dataset. The video dataset was composed of 416 videos. For training DeepSORT, we prepared the tumor region by cutting it out from the video data. Four-fifths, one-tenth, the remaining data of the image groups were used for training, validation, and test of our system, respectively.

TABLE II. THE DETAILS OF THE VIDEO DATASET

Diagnosis	# of videos
Hepatic Cell Carcinoma	32
Metastatic Cancer	46
Hemangioma	162
Cyst	151
Other Tumor	26
Total	416

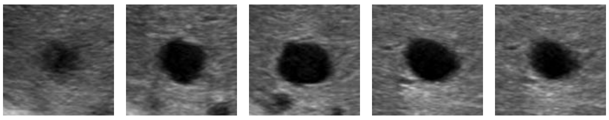


Figure 4. A set of tumor image data.

V. EXPERIMENTS AND ANALYSIS

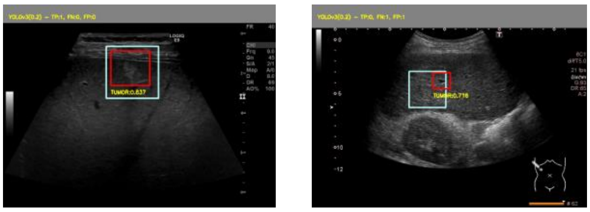
A. *Evaluation of Liver Tumor Detection Process*

For the liver tumor detection process, we used a YOLOv3 model pre-trained on ImageNet [11] and retrained on the AMED dataset. The training was conducted for 100 epochs. We finally adopted the model at the 71st epoch, when the loss on the validation data was lowest. As for evaluation metrics for our system, we calculated recall, precision, and F-measure. We regarded a detected region as a correct one if its IoU was equal to or more than 0.2, considering the ROIs provided from the AMED had margins around tumors of interest.

Experimental results showed that the recall, the precision, and F-measure were 0.9261, 0.9050, and 0.9154, respectively. The detection results are shown in Fig. 5. We can see in Fig. 5(b) that a false positive was observed in the blood vessel's cross-section or circular organ, which had a similar appearance to cysts.

B. *Evaluation of Liver Tumor Tracking*

For liver tumor tracking, we used a DeepSORT model pre-trained by MARS [12] and retrained on the AMED dataset. We trained the model for 520,000 steps and finally adopted the model at the 67,320th step, whose precision for the validation data was 0.9521. We evaluated a detected region in the same manner to the experiment described in the previous section. If a positive detection was found in one frame of grouped clusters, the cluster was regarded as positive. We adapted our method to the evaluation data. As a result, we were able to track positive clusters in 36 out of 37 evaluation data using the proposed method. In the evaluation example, two types of results are shown. The first is the evaluation data that shows the same tumor only once. This result mainly confirms that the tumor is being tracked. The second is the evaluation data that shows the same tumor multiple times. This result is used to check whether the reappeared tumor can be retraced as the same tumor.

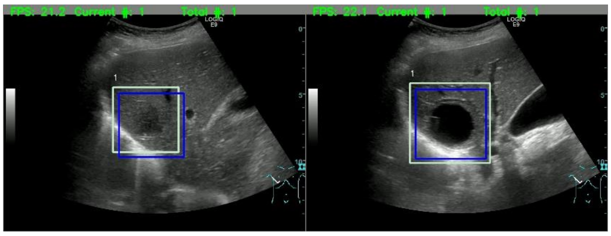


(a) Successful detection. (b) Failure detection.

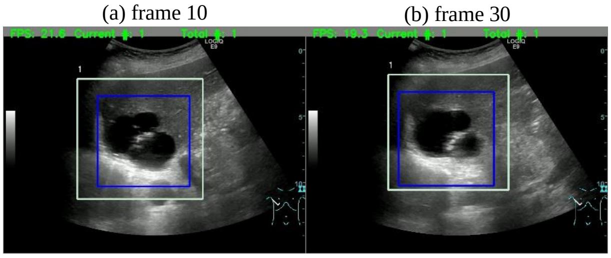
Figure 5. Examples of the detection result (The light blue and the red squares indicate lesions and the detected regions, respectively.).

1) *In case the same tumor shows up only once*

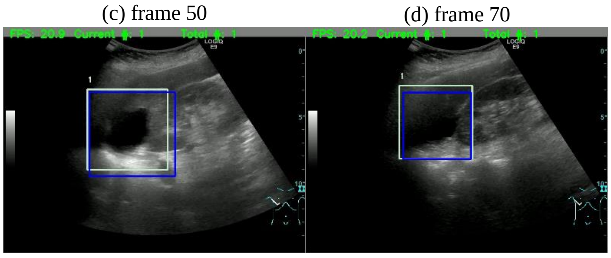
Fig. 6 and Fig. 7 show examples of successful tracking. Fig. 6 shows that the tumor in the center of the image can be tracked as the same tumor instance. This means that DeepSORT was effective in tracking tumors considering the overlap ratio.



(a) frame 10 (b) frame 30



(c) frame 50 (d) frame 70



(e) frame 90 (f) frame 110

Figure 6. Example of successful tracking (The blue box represents the lesion. The other colored boxes and the numbers in the upper left corner of the boxes indicate the tracking box and the cluster number.).

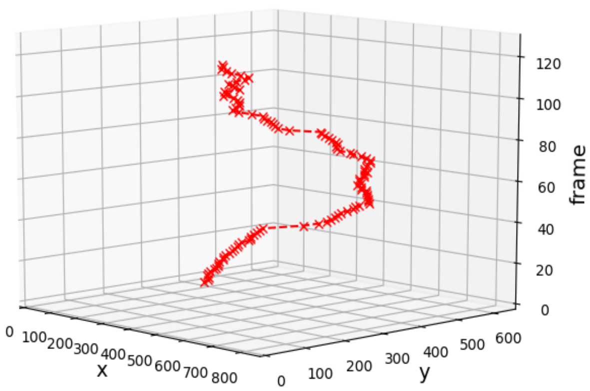
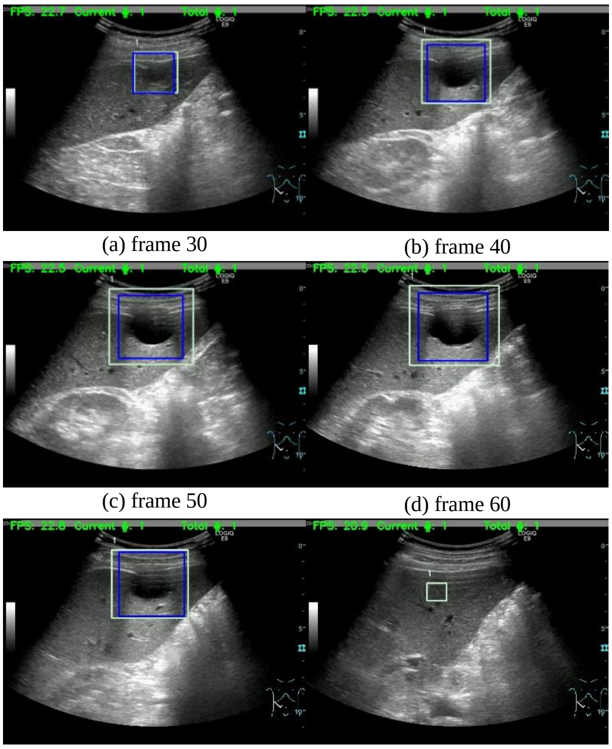


Figure 7. Center of gravity of each cluster for successful tracking (The red and blue point clusters indicate the positive and negative clusters, respectively.).

Fig. 8 and Fig. 9 show examples of failed tracking. Fig. 8(a) to Fig. 8(e) are successful ones, whereas Fig. 8(f) is failed one. We can see from these results that the false-positive blood vessel region was also tracked as the same tumor. This was probably because the DeepSORT-based CNN model extracts image features without considering the ROI size.



(a) frame 30 (b) frame 40

(c) frame 50 (d) frame 60

(e) frame 70 (f) frame 80

Figure 8. Example of failure tracking (The blue box represents the lesion. The other colored boxes and the numbers in the upper left corner of the boxes indicate the tracking box and the cluster number.).

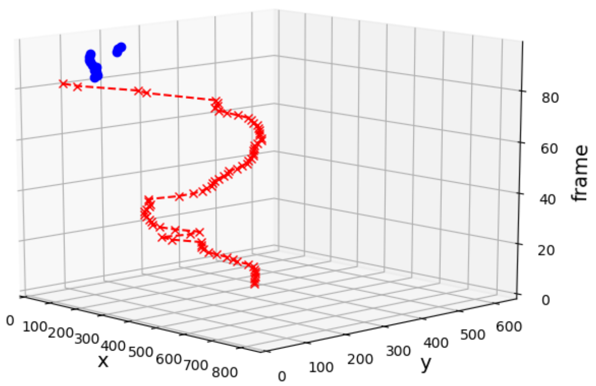
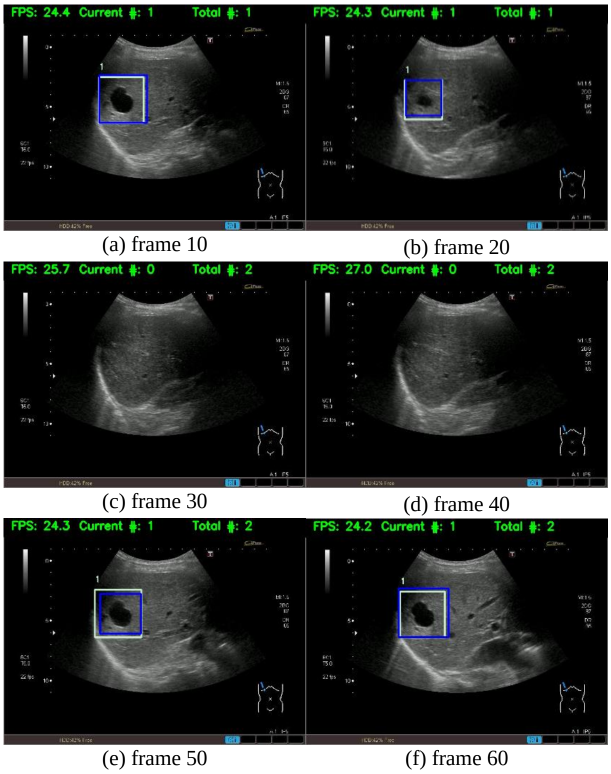


Figure 9. Center of gravity of each cluster for failure tracking (The red and blue point clusters indicate the positive and negative clusters, respectively.).

2) In case the same tumor shows up multiple times

Fig. 10 and Fig. 11 show examples of successful tracking. The tumor which was temporarily hidden by probe operation could be tracked as the same tumor instance. This may be because the process of the image feature extraction in DeepSORT could connect the tumor regions between distant frames.

Fig. 12 and Fig. 13 show an example of failed tracking. For the case of Fig. 13, we expected that the upper and the lower red point clusters are connected. The results, however, were not. This may be because its appearance when it disappeared was different from that when it reappeared.



(a) frame 10 (b) frame 20

(c) frame 30 (d) frame 40

(e) frame 50 (f) frame 60

Figure 10. Example of successful tracking (The blue box represents the lesion. The other colored boxes and the numbers in the upper left corner of the boxes indicate the tracking box and the cluster number.).

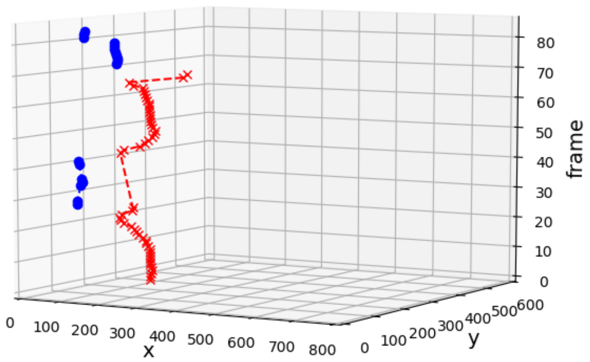
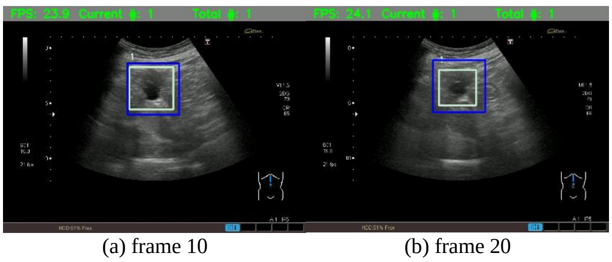
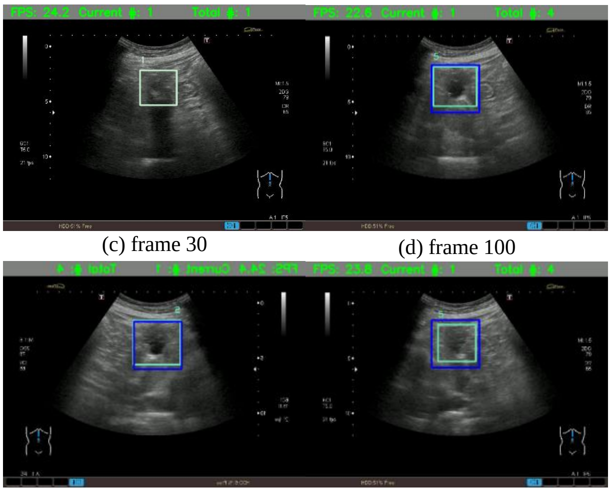


Figure 11. Center of gravity of each cluster for successful tracking (The red and blue point clusters indicate the positive and negative clusters, respectively.).



(a) frame 10 (b) frame 20



(c) frame 30 (d) frame 100

(e) frame 110 (f) frame 120

Figure 12. Example of failure tracking (The blue box represents the lesion. The other colored boxes and the numbers in the upper left corner of the boxes indicate the tracking box and the cluster number.).

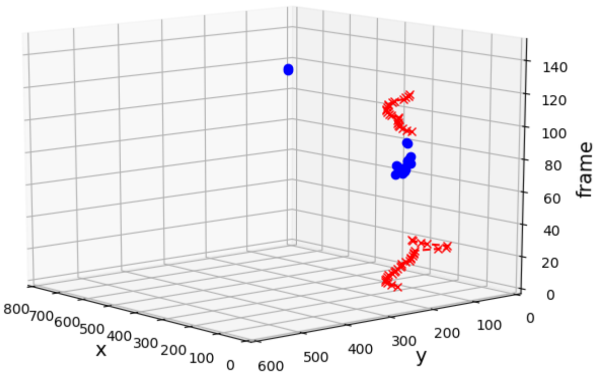


Figure 13. Center of gravity of each cluster for failure tracking (The red and blue point clusters indicate the positive and negative clusters, respectively.).

VI. Conclusion

We proposed a method of tumor tracking from an ultrasound video. Experimental results showed that our method could detect most of the lesions, and achieved F-measure of 0.9154 from a single image. In another experiment, we were able to track positive clusters in 36 out of 37 evaluation data using the proposed method.

Future work includes the accuracy improvement of tracking tumors that are temporarily hidden by probe manipulation. We would like to develop a method to take advantage of the most similar features of the tumor when it disappears and when it reappears.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Shoya Yamagishi, Keisuke Doman and Yoshito Mekada conducted the research, analyzed the data, and wrote the paper; Naoshi Nishida and Masatoshi Kudo conducted medical analysis; All authors had approved the final version.

Acknowledgment

Part of this research is supported by grants from the Japan Society for the Promotion of Science and the Japan Agency for Medical Research and Development (AMED).

References

[1] T. Nakashima, I. Tsutsumi, H. Takamia, K. Doman, Y. Mekada, N. Nishida, and M. Kudo, "A study on liver tumor detection from an ultrasound image using deep learning," in *Proc. International Workshop on Advanced Imaging Technology*, 2020.

[2] J. Redmon and A. Farhadi. (Apr. 2018). YOLOv3: An incremental improvement. arXiv. [Online]. Available: https://arxiv.org/pdf/1804.02767.pdf

[3] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Y. Fu, and A. C. Berg., "SSD: Single shot multibox detector," in *Proc. European Conference on Computer Vision*, 2016, pp. 21-37.

[4] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137-1149, June 2017.

[5] Z. Zhong and J. Deng, "Real-Time detection based on modified YOLO for herniated intervertebral discs," in *Proc. 4th International Conference on Intelligent Information Processing*, November 2019, pp. 466-470.

[6] N. Wojke, A. Bewley, and D. Paulus, "Simple online and realtime tracking with a deep association metric," in *Proc. IEEE International Conference on Image Processing*, Beijing, China, 2017, pp. 3645-3649.

[7] A. Lemay. (Oct. 2019). Kidney recognition in CT using YOLOv3. arXiv. [Online]. Available: https://arxiv.org/pdf/1910.01268.pdf

[8] S. Xie, R. Girshick, P. Dollar, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, HI, USA, 2017, pp. 5987-5995.

[9] N. S. Punn, S. K. Sonbhadra, and S. Agarwal. (May 2020). Monitoring COVID-19 social distancing with person detection and tracking via fine-tuned YOLO v3 and Deepsort techniquesarXiv. [Online]. Available: https://arxiv.org/pdf/2005.01385.pdf

[10] A. Bewley, G. Zongyuan, F. Ramos, and B. Upcroft, "Simple online and realtime tracking," in *Proc. IEEE International Conference on Image Processing*, Phoenix, AZ, USA, 2016, pp. 3464-3468.

[11] J. Deng, W. Dong, R. Socher, L. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, Miami, FL, USA, 2009, pp. 248-255.

[12] L. Zheng, Z. Bie, Y. Sun, J. Wang, C. Su, S. Wang, and Q. Tian, "MARS: A video benchmark for large-scale person re-identification," in *Proc. European Conference on Computer Vision*, 2016.

Shoya Yamagishi received his B.S degree in engineering at the Chukyo University, Japan in 2020. He is currently studying for M.S. degree in engineering at the Chukyo University in Japan. His research is mainly the analysis of medical images using machine learning.

Keisuke Doman received his B.S. degree from the Department of Electrical and Electronic Engineering, his M.S. degree and Ph.D. from the Graduate School of Information Science at Nagoya University in 2007, 2009, and 2012, respectively. He is currently an Associate Professor at the School of Engineering, Chukyo University, Japan. His current research interests include the application of computer vision and pattern recognition to human activity support systems. He is a member of IEICE, IEEE, ACM, and INSTICC.

Yoshito Mekada received the B.S., M.S., and Ph.D. degrees in information engineering from Nagoya University, Japan in 1991, 1993, and 1996, respectively. He is currently a professor at the School of Engineering, Chukyo University, Japan. His current research interests include medical image recognition, computer vision, and their applications. He is a member of the Japanese Society of Medical Imaging Technology and the IEEE.

Naoshi Nishida received the M.D. from Osaka Medical School (Osaka Medical and Pharmaceutical University) in 1985 and Ph.D. degree in Medical Science from Kyoto University, Japan in 1993. He is currently a professor at the Department of Gastroenterology and Hepatology, Kindai University Faculty of Medicine, Japan. His current research interests include molecular carcinogenesis, immuno-oncology, and diagnosis and treatment of liver cancer. He is a member of the Japanese Society of Gastroenterology, the Japanese Society of Hepatology and the American Gastroenterological Association.

Masatoshi Kudo received the M.D. and Ph.D. degrees in Medical Science from Kyoto University, Japan in 1978 and 1986. He is currently a professor and chairman at the Department of Gastroenterology and Hepatology, Kindai University Faculty of Medicine, Japan. His current research interests include diagnosis and treatment of liver cancer. He is a member of the Japan Society of Hepatology the Japanese Society of Gastroenterology and the American Association for the Study of Liver Diseases.