

Received 8 April 2024, accepted 3 June 2024, date of publication 6 June 2024, date of current version 14 June 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3410323

## RESEARCH ARTICLE

# Development of a Highlighting System for Surgical Instruments in Total Knee Arthroplasty

**RYUSEI KASAI<sup>ID</sup>, (Student Member, IEEE), AND KOUKI NAGAMUNE<sup>ID</sup>, (Senior Member, IEEE)**

Graduate School Engineering, University of Fukui, Fukui 910-8507, Japan

Corresponding author: Ryusei Kasai (mf230099@g.u-fukui.ac.jp)

This work was supported by Teijin Nakashima Medical Company.

**ABSTRACT** Currently, there is a shortage of nurses worldwide. In total knee arthroplasty, many surgical instruments of similar shapes and sizes are available. As a result, medical accidents have occurred due to errors in the selection of these instruments. To solve these problems, several surgical instrument detection systems using object detection methods have been developed. However, these studies do not mention how to display the detection results. It is inefficient for scrub nurses to look at the display to check the detection results, and mistakes can occur. Therefore, in this study, we have developed a system that detects surgical instruments and uses the detection results to highlight the instruments using a projector like projection mapping. This study also describes a method for creating training data and an inference method to reduce the additional training time required when the number of detection targets is increased. In the experiments, we verify the effectiveness of the calibration system for the camera and projector. In addition, object detection is performed on a surgical instrument to verify the accuracy of the system. The results showed that the error after calibration is 0.15 cm. The IoU and F-measure of object detection are 0.911 and 0.958, respectively, indicating the effectiveness of the system.

**INDEX TERMS** Machine learning, medical robotics, object detection.

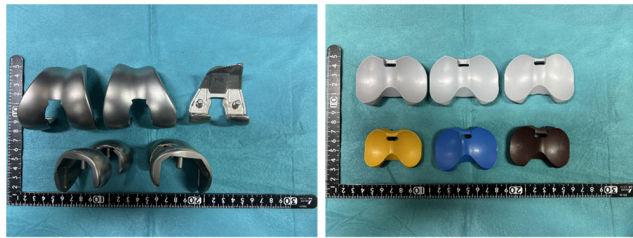
## I. INTRODUCTION

Currently, the shortage of nurses is processing all over the world and is becoming a major problem [1]. According to the WHO, there is a shortage of 5.9 million nurses worldwide [2]. From this, it is thought that there will be a similar or even greater shortage of scrub nurses, who are said to have a greater burden than regular nurses. Scrub nurses are nurses who prepare surgical instruments and hand them to surgeons during surgery [3]. Among these tasks, handing surgical instruments to the surgeon is very important task for performing surgery smoothly and safely. Scrub nurses need to memorize the names and shapes of all surgical instruments, which is a major burden. There are more than 100 types of surgical instruments to be memorized for complicated operations [4].

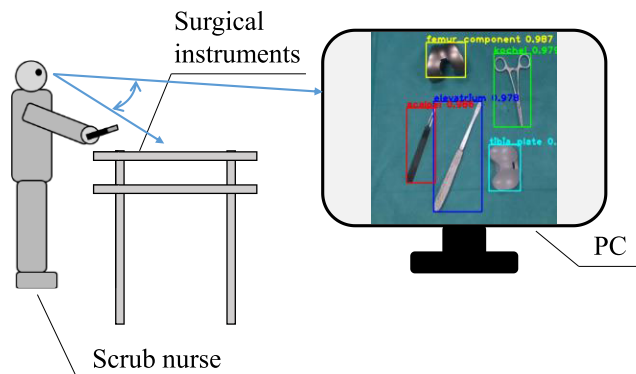
The associate editor coordinating the review of this manuscript and approving it for publication was Jingang Jiang<sup>ID</sup>.

Total Knee Arthroplasty (TKA) is a surgical procedure to replace a damaged knee caused by knee osteoarthritis or rheumatism with an artificial joint. TKA is expected to relieve knee joint pain and improve walking function. The following is a brief flow of TKA [5]. First, cut the skin on the front of the knee by 6-12 cm to expose the knee joint [6]. Next, the damaged bone is removed. Finally, the artificial joint is inserted and fixed. A large number of surgical instruments with similar shapes and sizes are used in TKA. Fig. 1 shows the implant used in the TKA. Medical accidents have occurred due to the incorrect selection of these implants, with a total of 169 cases of incorrect implant selection among 12,167 TKA and 10,680 total hip arthroplasty (THA) procedures performed in the United States from 2012 to 2017 [7]. As a result of this medical accident, it has been reported that patients have complained of knee pain and difficulty walking.

For these reasons, reducing the burden on scrub nurses and reducing the occurrence of medical accidents in TKA



**FIGURE 1.** Implants used in TKA. Femoral component and tibial polyethylene plate.



**FIGURE 2.** Conventional method of displaying detected objects. Inefficient and prone to errors because of the need to compare the PC and the surgical instrument.

are extremely important issues. Therefore, in this study, we developed a system that navigates scrub nurses to hand surgical instruments. Several studies have focused on surgical instrument detection [8], [9], [10], [11], [12]. Although these studies focused on detecting surgical instruments using Convolutional Neural Networks (CNN), they did not mention how to inform the position and name of detected surgical instruments to surgical nurses. However, when using the system in an actual operating room, how to inform the position and names of surgical instruments to scrub nurses is a very important issue. Generally, an object whose position and name are detected by object detection is displayed on a PC or other display as shown in Fig. 2, where the object is surrounded by a rectangle. In such a system, the scrub nurse must compare the rectangle on the PC or other display with the actual position of the surgical instruments, which is very inefficient and may cause errors due to a mistake in vision.

In addition, it is often decided to perform a surgery, such as an emergency surgery, the day before or just before the surgery is to be performed. Additionally, surgical instruments to be used may be added at the last minute. However, it takes a very long time to create trained files for detecting surgical instruments, and it may not be possible to start learning all the surgical instruments in time after the surgery is scheduled. As described above, with current science and technology, there is a problem with the method of informing the inference results to scrub nurses after detecting surgical instruments

using CNN, which limits its use in clinical practice. In addition, machine learning for object detection requires a lot of time, which is a limitation in emergency surgeries or when surgical instruments to be used are added urgently. Therefore, in this study, we created multiple files that had been trained in advance, selected training files to be used depending on the surgical instruments to be used, and used these files for inference. By adopting this method, even if a new surgical instrument that has not been trained is used, only that surgical instrument needs to be trained, so the number of detection targets can be increased with a low learning cost. This method is an algorithm that performs multiple inferences in parallel and compares the confidence in each prediction region from each inference.

The main contributions of this research are as follows. In this study, we developed an object detection system that detects surgical instruments used in TKA and allows scrub nurses to know the detection results without looking at a display such as a PC. In object detection, we propose an inference algorithm to reduce the learning time required for additional detection targets so that the system can handle additional surgical instruments immediately before surgery. In addition, as a display method, we propose a system that uses a projector to surround the actual surgical instruments like a projection mapping to inform the scrub nurse of the detection results. These new technologies make it possible to visually communicate the detection results of surgical instruments in the operating room, and by making it possible to add detection targets in a short time, providing a solution to problems such as the shortage of nurses and medical accidents. In addition, we verify the accuracy of object detection of the surgical instruments and the accuracy of the system for highlighting the surgical instruments using a projector in this paper.

## II. RELATED WORKS

Systems to support surgery have been actively developed, and “da Vinci” and “ROBODOC” are in use worldwide [13], [14]. Furthermore, systems to support scrub nurses are being developed [4], [15], [16]. These systems support the progress of medicine and are considered to be in great demand.

In recent years, there has been active development of systems combining surgical support systems and CNN [8], [9], [10], [11], [12]. This is due to advances in CNN-based object detection techniques such as Fast R-CNN, Faster R-CNN, You Only Look Once (YOLO), and Single Shot MultiBox Detector (SSD) [17], [18], [19], [20]. Nakano et al. developed a system that uses Faster R-CNN to detect the position and type of five types of surgical instruments commonly used in surgery [8]. Lee et al. investigated the recognition rate of 14 types of surgical instruments using Faster R-CNN, Mask R-CNN, and SSD-based methods [10]. These studies focused on the detection of surgical instruments and did not specifically mention how to display the detection results. However, when used in an actual operating room environment, the method for displaying detection results is a very important

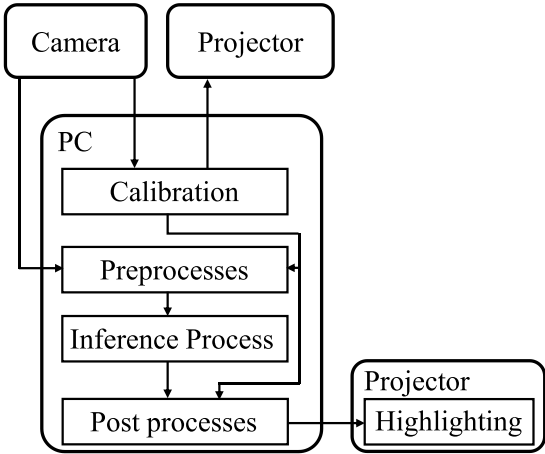


FIGURE 3. Simple flow of highlighting system.

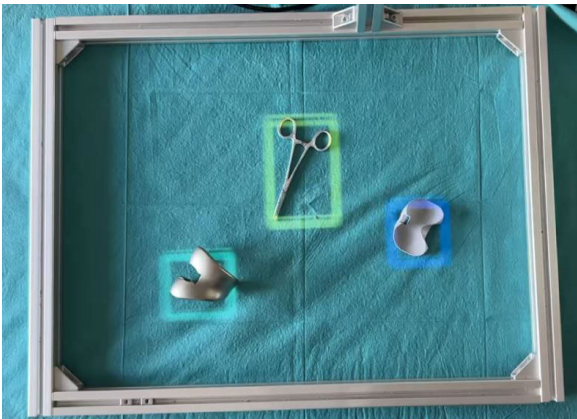


FIGURE 4. Highlighting surgical instruments by the system. Kocher is surrounded by green, femoral component is surrounded by light blue, and tibial polyethylene plate is surrounded by blue.

factor. Therefore, in this study, we developed a system that focuses not only on CNN-based object detection methods but also on the display method of detection results.

### III. METHOD

#### A. SYSTEM OVERVIEW

Fig. 3 shows a simple flow of the entire system. First, a camera placed at a 60-75 cm height is used to capture surgical instrument images. The captured image is pre-processed, and object detection is performed on the pre-processed image. YOLOv5, an object detection method, is used for object detection. Multiple inferences are performed in parallel, and the detection results are obtained by comparing the confidence levels. Finally, the projector placed at a height of 85 cm surrounds the surgical instrument with a rectangle. Fig. 4 shows how surgical instruments are highlighted using this system. The details of each of them are shown below.

#### B. DEVICES

Fig. 5 shows the developed system. I used one RealSense D415™ (Intel Co.) as the camera [21]. Table 1 shows the

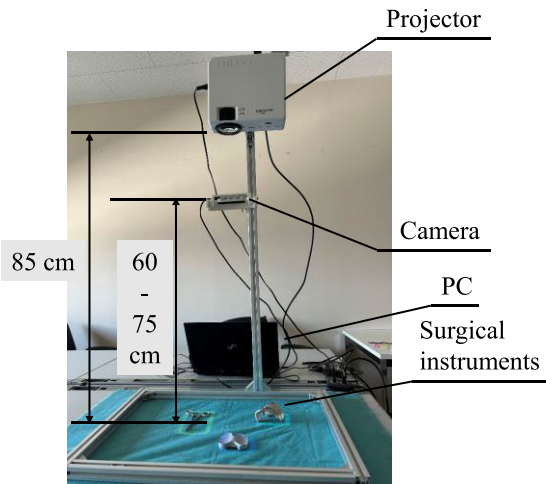


FIGURE 5. The developed system. It consists of a camera, PC, and projector.

TABLE 1. Specification of RealSense D415™.

Method of RGB-D camera	Active IR stereo (Rolling shutter)
Field of view	69.4° × 42.5° × 77.0° (± 3°)
Minimum depth distance	0.3 m
Resolution, Frame rate (Depth)	Max 1280 × 720, Max 90 fps
Resolution, Frame rate (RGB)	Max 1920 × 1080, Max 30 fps
Maximum range	About 10 m

TABLE 2. Specification of the projector.

Resolution	1920 × 1080
Imaging system	LCD
Light source	LED
Recommended projection distance	1.25 – 4 m

TABLE 3. Specification of the PC.

Model	iiyama P15SM1-A
CPU	Intel® Core™ i7-4810MQ CPU @ 2.80 GHz
Memory	32 GB
OS	Ubuntu 18.04.6 LTS

specifications of RealSenseD415™. ROVOMKO MKO26 was used as the projector. Table 2 shows the specifications of the projector. Table 3 shows the specifications of the PC. This PC was used for camera and projector calibration, inference, and visualization. Machine learning for object detection requires a high specification GPU. Therefore, we used Google Colaboratory for machine learning, which allows the use of high-performance GPU online.

In this study, YOLO was used for inference [19]. YOLO is a CNN-based object detection algorithm. Object detection algorithms can be largely divided into one-stage and two-stage types. One-stage object detection algorithms such as YOLO and SSD use the same network for object region inference and classification. YOLO and SSD have excellent real-time performance due to their simple structure of dividing the input image into arbitrary grids and performing object detection on each grid [19], [20]. Multiple versions of YOLO have been released so far, and Yan et al. compared YOLOv3, v4, and v5 and found that YOLOv5 has superior mAP and processing speed, and has a smaller model size [22]. For this reason, we used YOLOv5 among several versions of YOLO in this study. Furthermore, YOLOv5 has released models of different sizes (model s, model m, model l, and model x) with different detection accuracy and speeds [23].

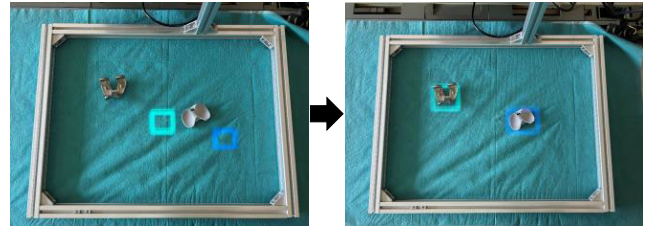
### C. CALIBRATION OF CAMERA AND PROJECTOR

Before using the system, it is necessary to obtain the positional relationship between the camera and projector and perform calibration to match the coordinate system. This calibration makes it possible to surround the correct position with a rectangle as shown in Fig. 6. Fig. 7 shows the calibration process. First, a rectangle is illuminated by the projector. The size of the rectangle is adjusted according to the positions of the projector and the camera so that all rectangles are captured. This rectangle is the range where detection and illumination are possible. Next, the position of the illuminated rectangle is obtained by image processing, and distortion is removed by performing a projective transformation based on the position of the rectangle. The size after the projective transformation is based on the size of the obtained rectangle. Finally, the coordinates of the rectangle on the projector and the coordinates of the rectangle on the camera are matched by projective transformation. 1 shows the projective transformation for transforming the coordinates  $(x, y)$  to the coordinates  $(x', y')$ . Through this process, a projection matrix is obtained. This is the calibration method for the camera and projector.

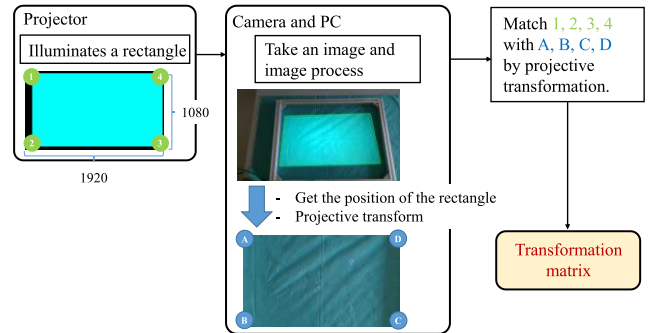
$$\begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} = \begin{pmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \quad (1)$$

### D. TRAINING DATA CREATION PROCESSING

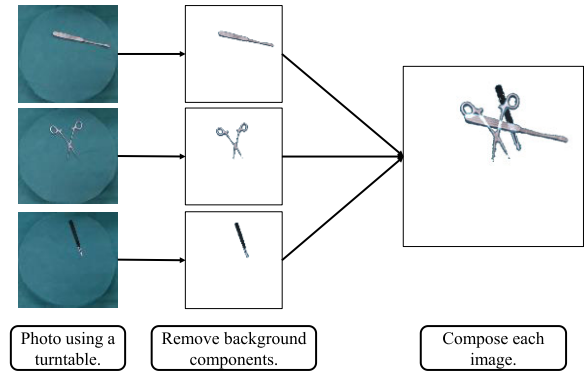
This section describes the creation of training data for machine learning. The method of creating training data is also different from the usual method because, in this research, multiple inferences are performed in parallel, instead of the usual inference method. In this study, one dataset is a combination of 1-3 surgical instruments. Each surgical instrument was captured separately using a turntable, and the background components are removed to create image data as an RGBA image. Then, they were combined into a single image dataset (Fig. 8). Each image data has annotation data indicating the



**FIGURE 6.** The state before and after calibration. Before calibration (left), the illuminated rectangle is misaligned. After calibration (right), the rectangle position is correct.



**FIGURE 7.** Calibration process.



**FIGURE 8.** Method of dataset creation. Remove the background component of the surgical instrument captured by using a turntable. Then, they are combined into a single image.

object position and this annotation data is also combined when the images are combined. The training was performed using the combined image data and standard objects (Fig. 9) as a single dataset. The standard objects were selected to be similar in color and shape to the surgical instruments to improve the detection accuracy. Standard object A (Fig. 9 (a)) is a 15 cm × 1.7 cm × 0.8 cm bar with a 9 cm section on one side covered in black. Standard Object B (Fig. 9 (b)) is a 5 cm × 5 cm × 1.5 cm white rectangle. Standard object C (Fig. 9 (c)) is an 8 cm × 7 cm × 0.6 cm black rectangle.

### E. INFERENCE PROCESSING

Fig. 10 shows the inference flow. This inference algorithm is an improvement of the previous authors' inference

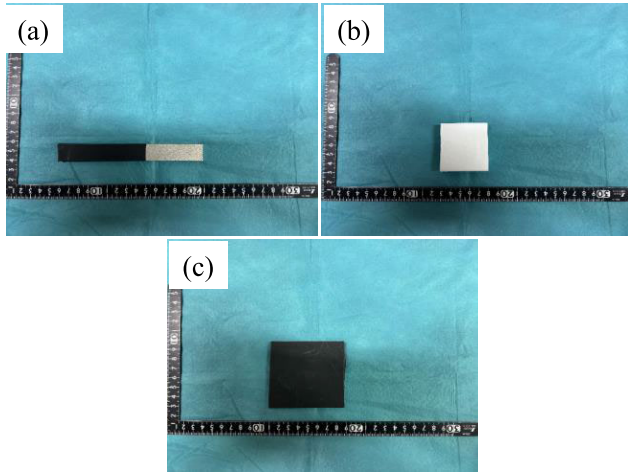


FIGURE 9. Standard objects.

algorithm [24]. Inference is performed using multiple trained files. First, the background component of the input image is removed to obtain an RGBA image. This image is input to all trained files, and multiple inferences are performed in parallel. Each inference outputs the predicted region of the object, its label name, and confidence level. Here, results that are detected as standard objects are excluded. The upper-left and lower-right coordinates of each predicted region are compared, and if the difference between them is less than a threshold, the prediction is for the same region. The inference results that are determined to be the same are compared in confidence level and the inference result with the highest confidence is used as the prediction for that region. Furthermore, the object position is obtained through image processing, compared with the inferred object position, and the one that is determined to match is taken as the final detection result.

With this inference algorithm, the number of detectable objects can be increased simply by adding weight files trained on the surgical instruments to be added, resulting in significant time savings.

#### F. VISUALIZATION PROCESSING

The object position and label name obtained through the inference are used for visualization. The position of this object is the coordinates on the image used in the inference, so it needs to be converted to the coordinates of the real world where the surgical instruments are located.

The projection matrix obtained in the calibration is applied to the coordinates on the image to obtain the coordinates in the real world. By changing the color of the rectangle according to the label name, the user can easily identify “which surgical instrument is at which position”.

#### G. EVALUATION INDEX

In this study, Intersection of Union (IoU), Precision, Recall, F-measure, and frames per second (fps) were used as evaluation metrics. IoU is the ratio of overlap between the correct

TABLE 4. Confusion matrix.

		Prediction	
		Positive	Negative
Actual	Positive	TP	FN
	Negative	FP	TN

area of the object and the predicted area. IoU is calculated as

$$IoU = \frac{\text{Area of Overlap}}{\text{Area of Union}} \quad (2)$$

IoU takes values between 0 and 1, and IoU 1 indicates that the correct area is predicted without over- or under-prediction. In addition, evaluation is performed using true positive (TP), false positive (FP), false negative (FN), and true negative (TN). The confusion matrix is shown in Table 4. In this study, if the IoU between the correct region and the predicted region of the object is 0.5 or more, it is considered TP, and if it is less than 0.5, it is considered FP. FN occurs when a surgical instrument that should be detected is missed. Since there are countless TN in object detection, they are not used. Precision, Recall, and F-measure are defined as

$$\text{Precision} = \frac{TP}{TP + FP} = \frac{TP}{\text{All detection}} \quad (3)$$

$$\text{Recall} = \frac{TP}{TP + FN} = \frac{TP}{\text{All ground truths}} \quad (4)$$

$$F - \text{measure} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (5)$$

Precision is the ratio of correctly detected areas among the predicted areas, and Recall is the ratio of correctly predicted areas among the areas to be detected. F-measure is the harmonic mean of Precision and Recall and can be evaluated comprehensively. In addition, fps is calculated from the time taken to process a certain number of images and is expressed as

$$\text{fps} = \frac{\text{Number of images processed}}{\text{Required time [s]}} \quad (6)$$

## IV. EXPERIMENTS

### A. VERIFICATION OF CALIBRATION ACCURACY

In 1<sup>st</sup> experiment, a basic experiment was conducted to verify the accuracy of the calibration of the camera and projector. The experimental method is described below. First, nine points as shown in Fig. 11 were prepared. Then, the calibrated system is used to illuminate a circle at each point captured by the camera. The distances between point 1 and the other points are measured and the error from the true value is measured: 8 cm between points 1-2, 1-3, 1-4 and 1-5, 20 cm between points 1-6, 1-7, 1-8 and 1-9. The height of the camera was changed to 60 cm, 70 cm, and 80 cm. The height of the camera is represented by H in Fig. 12.  $\theta$  in Fig. 12 represents the angle of the camera, which was fixed at 10° in this experiment.

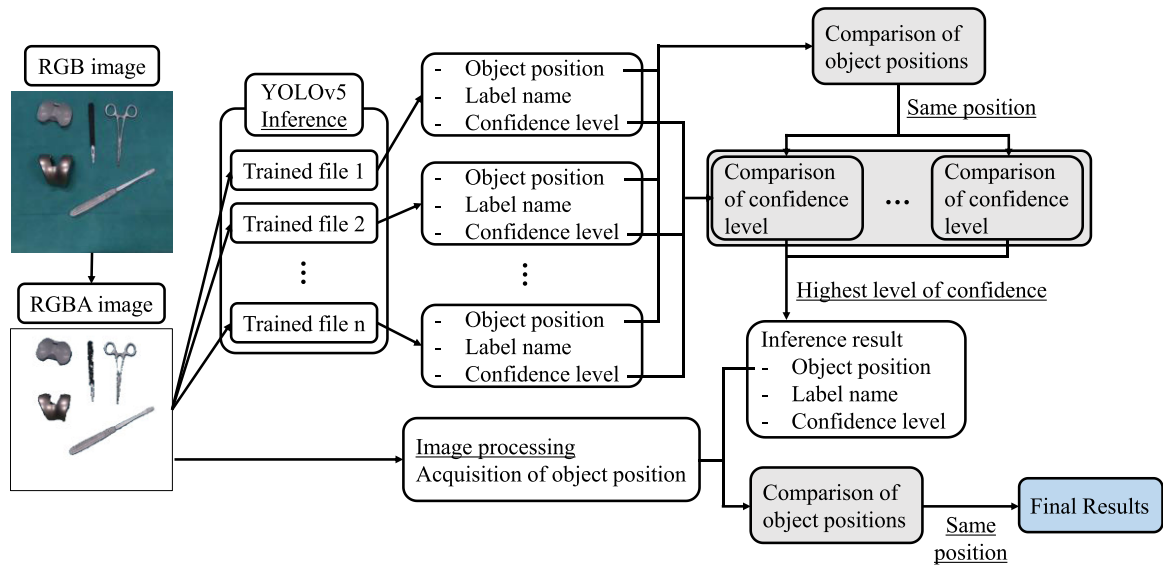
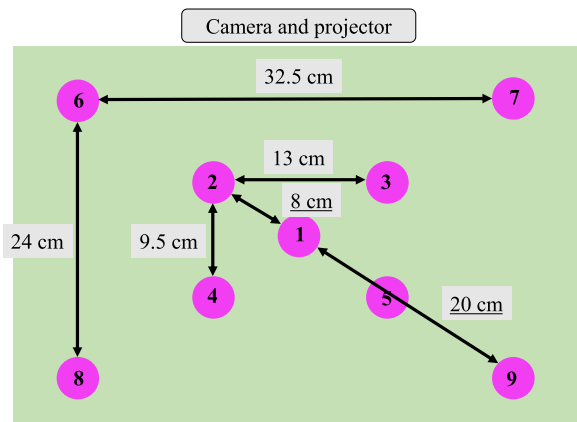
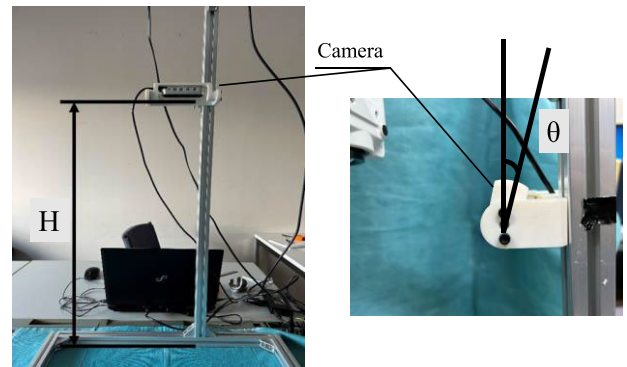


FIGURE 10. Flow of inference.

FIGURE 11. Nine points used in 1<sup>st</sup> experiment.FIGURE 12. Environment of 1<sup>st</sup> experiment, where H is the height of the camera and  $\theta$  is the angle of the camera.

## B. OBJECT DETECTION FOR SURGICAL INSTRUMENTS

In 2<sup>nd</sup> experiment, object detection for surgical instruments was performed. Elevators, kocher, and scalpel, which are widely used in general surgery, and femoral component and tibial polyethylene plate, which are used in TKA, were used as detection targets. Each surgical instrument is shown in Fig. 13. In addition, the combinations of training data used in this experiment and the number of labels for each surgical instrument are shown in Table 5. In Table 5, the femoral component was trained separately for the front, side, and back of the femoral component, because the shape of the femoral component is very different in each orientation. Three types of femoral components with different sizes and orientations were used in the study (Model number: 3L left, 2L right, and M right). Tibial polyethylene plates are divided into two groups because of their respective colors. Tibial polyethylene plate 1 is a gray plate (Model number: 3L 05, 3L 09, 3L 11) and tibial polyethylene plate 2 is other colors (Model number: L 05, M 05, S 05). This experiment was trained

using YOLOv5 (model m), with a batch size of 16 and 100 epochs. The time required for training was 1.175 hours for Group 1, 0.804 hours for Group 2, 0.721 hours for Group 3, and 0.803 hours for Group 4. The GPU assigned by Google Colaboratory at this time was a Tesla T4. A total of 100 images showing all surgical instruments were used as test data. An example of test data is shown in Fig. 14. Each surgical instrument was placed in various orientations without overlapping. Because the kocher was made like scissors, it was placed open at various angles. IoU, Precision, Recall, and F-measure were used as evaluation indicators. Additionally, fps was calculated from the number of images processed per minute.

## V. RESULTS

### A. VERIFICATION OF CALIBRATION ACCURACY

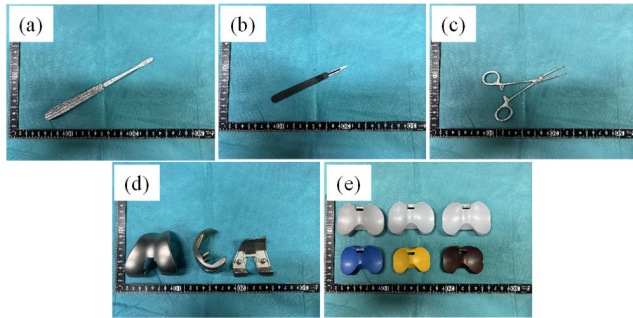
The results of 1<sup>st</sup> experiment are shown in Table 6. Table 6 shows the distances between 1-2, 1-3, 1-4, 1-5, 1-6, 1-7, 1-8, and 1-9 when the camera height is 60 cm, 70 cm, and 80 cm. It also shows the error from the true value.

**TABLE 5.** Combination of surgical instruments used for machine learning and number of labels.

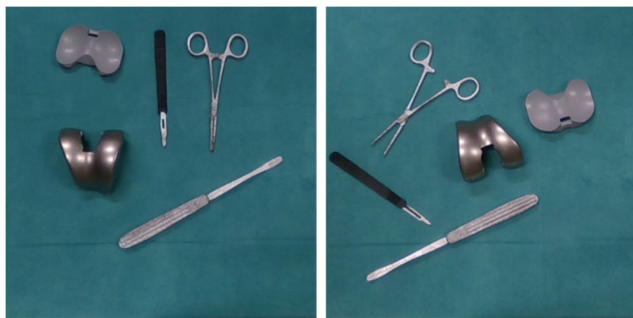
Group	Surgical instrument 1 Number of labels (train, validation)	Surgical instrument 2 Number of labels (train, validation)	Surgical instrument 3 Number of labels (train, validation)	Standard objects Number of labels (train, validation)
1	Elevators (640, 160)	Scalpel (640, 160)	Kocher (640, 160)	Standard objects (240, 60)
2	Surface of femoral component (320, 80)	Tibial polyethylene plate 1 (320, 80)		Standard objects (240, 60)
3	Side of femoral component (320, 80)			Standard objects (240, 60)
4	Back of femoral component (320, 80)	Tibial polyethylene plate 2 (320, 80)		Standard objects (240, 60)

**TABLE 6.** Results of 1<sup>st</sup> experiment, showing distances between 1-2, 1-3, 1-4, 1-5, 1-6, 1-7, 1-8, and 1-9 and their errors from the true values.

True value [cm]		8				20				
Height [cm]	Between 1-x	1-2	1-3	1-4	1-5	1-6	1-7	1-8	1-9	Ave.
60	Measured value	8.1	8.1	8	8.2	20.1	20.5	20	20.2	
	error	0.1	0.1	0	0.2	0.1	0.5	0	0.2	0.15
70	Measured value	7.9	7.8	7.9	7.8	20	19.8	20.2	20.4	
	error	0.1	0.2	0.1	0.2	0	0.2	0.2	0.4	0.175
80	Measured value	8.4	8.6	8.2	7.5	20.4	20.6	20.2	19.6	
	error	0.4	0.6	0.2	0.5	0.4	0.6	0.2	0.4	0.4125



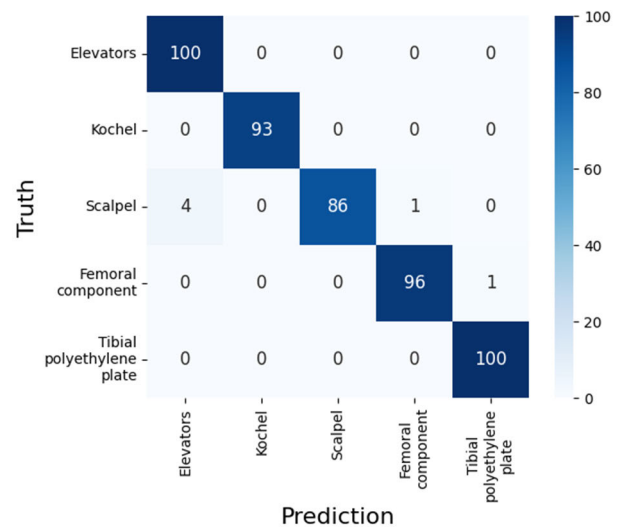
**FIGURE 13.** Surgical instruments used in 2<sup>nd</sup> experiment. (a) Elevators, (b) Scalpel, (c) Kocher, (d) Femoral component, (e) Tibial polyethylene plate.



**FIGURE 14.** Test data for 2<sup>nd</sup> experiment.

## B. OBJECT DETECTION FOR SURGICAL INSTRUMENTS

Table 7 shows the IoU, Precision, Recall, and F-measure for each surgical instrument. It took 283 seconds to process



**FIGURE 15.** Confusion matrix of the inference results in 2<sup>nd</sup> experiment.

100 images. This means that the fps is approximately 0.353. Fig. 15 shows the confusion matrix of the inference results. Fig. 16 shows the detection results.

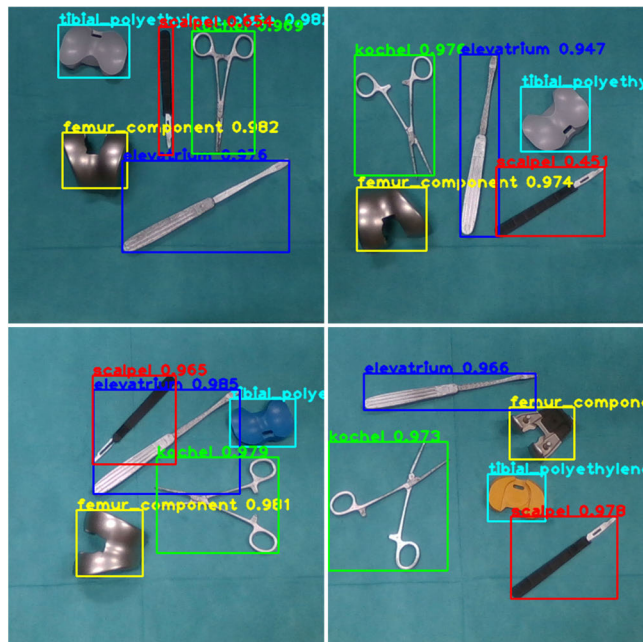
## VI. DISCUSSION

### A. VERIFICATION OF CALIBRATION ACCURACY

The average error when the camera height was 60 cm, 70 cm, and 80 cm was 0.15 cm, 0.175 cm, and 0.4125 cm, respectively. Although the errors increased as the height of the

**TABLE 7. Results of object detection for surgical instruments.**

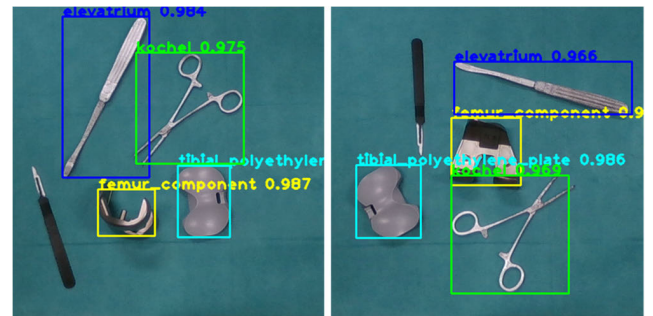
Surgical instrument	Elevators	Kocher	Scalpel	Femoral component	Tibial polyethylene plat	Ave.
IoU	0.955	0.932	0.895	0.819	0.955	0.911
Precision	0.952	0.989	0.989	0.960	0.952	0.968
Recall	1.000	0.930	0.860	0.960	1.000	0.950
F-measure	0.975	0.959	0.920	0.960	0.975	0.958

**FIGURE 16. Result of object detection in 2<sup>nd</sup> experiment.**

camera increased, they were not so large that they would be difficult to distinguish when the images were highlighted. Therefore, it can be said that the calibration of the system was performed correctly. The reason that the error increased as the height increased is thought to be because the quality of the image captured by the camera deteriorates. Since the results are particularly good at 60 cm and 70 cm, the appropriate height of the camera in this system is set to 60-70 cm.

## B. OBJECT DETECTION FOR SURGICAL INSTRUMENTS

In general, an IoU score of 0.8 or higher is considered sufficient for detecting the correct region, and since the average IoU score was 0.911, it can be said that the correct region was detected for all surgical instruments. Next, we focus on the F-measure, which is the harmonic mean of Precision and Recall, and if F-measure also shows a score of 0.8 or higher, it is said to be able to detect the correct region adequately. In this experiment, the average of F-measure was 0.958, indicating that the detection was sufficiently accurate. However, the scalpel's Recall was slightly lower than that of other surgical instruments. This is because there were some cases of missed detection as shown in Fig. 17. To address this problem, it is necessary to increase the number of training data. However,

**FIGURE 17. False detection for scalpel.**

overall, it shows a sufficient score, and it can be said that the dataset creation method and inference method used in this study are effective methods.

Furthermore, the fps score was 0.353. Although it cannot be said that the detection is completely real-time, the score is sufficient because the position of the surgical instrument does not change frequently during surgery.

The following is a comparison of the results with previous studies. However, there are few studies that perform object detection and evaluation using CNN for multiple surgical instruments, and sufficient comparisons cannot be made. In this paper, we compare the results of Nakano et al. using Faster R-CNN [8] and Bajraktari et al. using YOLOv5 and Mask R-CNN [25] among the studies that detect multiple surgical instruments. Nakano et al. detected five types of surgical instruments and the average F-measure was 0.931. Bajraktari et al. detected six types of surgical instruments, and the average F-measure was 0.830 when using YOLOv5 and 0.795 when using Mask R-CNN. Since our average F-measure was 0.958, it can be said that the dataset creation method and inference method of this study are effective.

From the above, we conclude that the inference in this study works well.

## VII. CONCLUSION

The purpose of this study was to detect surgical instruments used in TKA and highlight the detection results. We also aimed to shorten the learning time required when increasing the number of detection targets. For this purpose, we developed an inference algorithm that trains on multiple datasets and compares confidence levels. We also developed a system that surrounds the detection results like projection mapping using a projector. In order to test the effectiveness of the system, we verified the accuracy of the calibration of the

camera and projector and the accuracy of object detection in experiments. In experiments of camera and projector calibration, it was confirmed that the calibration error was less than 0.4125 cm. In addition, the accuracy of object detection was 0.958 in F-measure. From the two experiments, we can conclude that the system developed in this study is effective. In the future, we will increase the number of cameras and use depth information to improve inference accuracy and component size estimation. We will also develop a system that can be used more effectively in operating rooms by combining this system with voice recognition technology. Furthermore, we will conduct experiments in an actual operating room environment to evaluate the detection results. The highlighting system can be applied to various fields other than surgical instruments and will contribute to the development of the industry.

## ACKNOWLEDGMENT

The authors would like to thank Teijin Nakashima Medical Company for collaboration on this study.

## REFERENCES

- [1] M. Marć, A. Bartosiewicz, J. Burzyńska, Z. Chmiel, and P. Januszewicz, "A nursing shortage—A prospect of global and local policies," *Int. Nurs. Rev.*, vol. 66, no. 1, pp. 9–16, Mar. 2019.
- [2] *State of the World's Nursing 2020: Investing in Education, Jobs and Leadership*, World Health Org. (WHO), Geneva, Switzerland, 2020.
- [3] L. Mitchell and R. Flin, "Non-technical skills of the operating theatre scrub nurse: Literature review," *J. Adv. Nursing*, vol. 63, no. 1, pp. 15–24, Jul. 2008.
- [4] C. Perez-Vidal, E. Carpintero, N. Garcia-Aracil, J. M. Sabater-Navarro, J. M. Azorin, A. Candela, and E. Fernandez, "Steps in the development of a robotic scrub nurse," *Robot. Auto. Syst.*, vol. 60, no. 6, pp. 901–911, Jun. 2012.
- [5] T. Kamiya and S. Kawaguchi, "A development of force distribution measurement system in total knee arthroplasty and reporting of the PS type measurement," in *Proc. 29th Fuzzy Syst. Symp.*, 2013, pp. 238–241.
- [6] R. Vaishya, V. Vijay, D. Demesugh, and A. K. Agarwal, "Surgical approaches for total knee arthroplasty," *J. Clin. Orthopaedics Trauma*, vol. 7, no. 2, pp. 71–79, 2016.
- [7] M. P. Ast, D. J. Mayman, M. P. Bostrom, A. G. D. Valle, and S. B. Haas, "Can we avoid implant-selection errors in total joint arthroplasty?" *Clin. Orthopaedics Rel. Res.*, vol. 477, no. 1, pp. 130–133, 2019.
- [8] A. Nakano and K. Nagamune, "A development of robotic scrub nurse system-detection for surgical instruments using faster region-based convolutional neural network," *J. Adv. Comput. Intell. Intell. Informat.*, vol. 26, no. 1, pp. 74–82, Jan. 2022.
- [9] Z. Zhao, T. Cai, F. Chang, and X. Cheng, "Real-time surgical instrument detection in robot-assisted surgery using a convolutional neural network cascade," *Healthcare Technol. Lett.*, vol. 6, no. 6, pp. 275–279, Dec. 2019.
- [10] J.-D. Lee, J.-C. Chien, Y.-T. Hsu, and C.-T. Wu, "Automatic surgical instrument recognition—A case of comparison study between the faster R-CNN, mask R-CNN, and single-shot multi-box detectors," *Appl. Sci.*, vol. 11, no. 17, p. 8097, Aug. 2021.
- [11] T. Cai and Z. Zhao, "Convolutional neural network-based surgical instrument detection," *Technol. Health Care*, vol. 28, pp. 81–88, Jun. 2020.
- [12] S. M. Cho, Y.-G. Kim, J. Jeong, I. Kim, H.-J. Lee, and N. Kim, "Automatic tip detection of surgical instruments in biportal endoscopic spine surgery," *Comput. Biol. Med.*, vol. 133, Jun. 2021, Art. no. 104384.
- [13] M. Hoeckelmann, I. J. Rudas, P. Fiorini, F. Kirchner, and T. Haidegger, "Current capabilities and development potential in surgical robotics," *Int. J. Adv. Robotic Syst.*, vol. 12, no. 5, p. 61, May 2015.
- [14] P. Koukourikis and K. H. Rha, "Robotic surgical systems in urology: What is currently available?" *Investigative Clin. Urol.*, vol. 62, no. 1, p. 14, 2021.
- [15] L. Wagner, S. Kolb, C. Looschen, L. Bernhard, J. Fuchtmann, M. Berlet, J. Fottner, A. Knoll, and D. Wilhelm, "Versatile end effector for laparoscopic robotic scrub nurse," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 18, no. 9, pp. 1589–1600, May 2023.
- [16] M. G. Jacob, Y.-T. Li, and J. P. Wachs, "A gesture driven robotic scrub nurse," in *Proc. IEEE Int. Conf. Syst., Man, Cybern.*, Oct. 2011, pp. 2039–2044.
- [17] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1440–1448.
- [18] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. Int. Conf. Adv. Neural Inf. Process. Syst.*, vol. 28, 2015, pp. 91–99.
- [19] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 779–788.
- [20] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single shot multibox detector," in *Proc. 14th Eur. Conf. The Netherlands: Springer*, Oct. 2016, pp. 21–37.
- [21] L. Keselman, J. I. Woodfill, A. Grunnet-Jepsen, and A. Bhowmik, "Intel(R) real sense(TM) stereoscopic depth cameras," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 1267–1276.
- [22] B. Yan, P. Fan, X. Lei, Z. Liu, and F. Yang, "A real-time apple targets detection method for picking robot based on improved YOLOV5," *Remote Sens.*, vol. 13, no. 9, p. 1619, Apr. 2021.
- [23] M. Sozzi, S. Cantalamessa, A. Cogato, A. Kayad, and F. Marinello, "Automatic bunch detection in white grape varieties using YOLOV3, YOLOV4, and YOLOV5 deep learning algorithms," *Agronomy*, vol. 12, no. 2, p. 319, Jan. 2022.
- [24] R. Kasai and K. Nagamune, "A deep learning approach for surgical instruments detection system in total knee arthroplasty-automatic creation of training data and reduction of training time," *J. Adv. Comput. Intell. Intell. Informat.*, vol. 28, no. 1, pp. 150–158, Jan. 2024.
- [25] F. Bajraktari, K. Fleissner, and P. P. Pott, "A comparison of two CNN-based instrument detection approaches for automated surgical assistance systems," *Current Directions Biomed. Eng.*, vol. 9, no. 1, pp. 599–602, Sep. 2023.



**RYUSEI KASAI** (Student Member, IEEE) received the B.E. degree in mechanical and system engineering from the University of Fukui, Fukui, Japan, in 2023, where he is currently pursuing the M.E. degree.

His research interests include machine learning and image processing.



**KOUKI NAGAMUNE** (Senior Member, IEEE) received the B.E., M.E., Ph.D. degrees in computer engineering from Himeji Institute of Technology, Hyogo, Japan, in 1999, 2001, and 2004, respectively.

He was a Researcher and a Lecturer with the Department of Orthopaedic Surgery, Graduate School of Medicine, Kobe University, Hyogo, from 2004 to 2005 and from 2005 to 2007, respectively, and a Lecturer with the Department of Human and Artificial Intelligent System, Graduate School of Engineering, University of Fukui, Fukui, Japan, from 2007 to 2010, where he has been an Associate Professor, since 2010. His research interests include computer-aided surgery, computer-aided diagnosis, signal processing, and fuzzy logic.

Dr. Nagamune is a member of Japanese Journal of Clinical Biomechanics and Japanese Society for Medical and Biological Engineering.

...