공학박사 학위논문

Deep Learning Approaches for Clinical Performance Improvement: Applications to Colonoscopic Diagnosis and Robotic Surgical Skill Assessment

임상술기 향상을 위한 딥러닝 기법 연구: 대장내시경 진단 및 로봇수술 술기 평가에 적용

2020 년 8 월

서울대학교 대학원

협동과정 바이오엔지니어링 전공

이 동 헌

Ph. D. Dissertation

Deep Learning Approaches for Clinical Performance Improvement: Applications to Colonoscopic Diagnosis and Robotic Surgical Skill Assessment

BY

DONGHEON LEE

AUGUST 2020

INTERDISCIPLINARY PROGRAM IN
BIOENGINEERING
THE GRADUATE SCHOOL
SEOUL NATIONAL UNIVERSITY

Abstract

Deep Learning Approaches for Clinical Performance Improvement: Applications to Colonoscopic Diagnosis and Robotic Surgical Skill Assessment

Dongheon Lee
Interdisciplinary Program in Bioengineering
The Graduate School
Seoul National University

This paper presents deep learning-based methods for improving performance of clinicians. Novel methods were applied to the following two clinical cases and the results were evaluated.

In the first study, a deep learning-based polyp classification algorithm for improving clinical performance of endoscopist during colonoscopy diagnosis was developed. Colonoscopy is the main method for diagnosing adenomatous polyp, which can multiply into a colorectal cancer and hyperplastic polyps. The classification algorithm was developed using convolutional neural network (CNN), trained with colorectal polyp images taken by a narrow-band imaging colonoscopy. The proposed method is built around an automatic machine learning (AutoML) which searches for the optimal architecture of CNN for colorectal polyp image classification and trains the weights of the architecture. In addition, gradient-weighted class activation mapping technique was used to overlay the probabilistic basis of the prediction result on the polyp location to aid the endoscopists visually. To verify the improvement in diagnostic performance, the efficacy of endoscopists with varying proficiency levels were compared with or without the aid of the proposed polyp classification algorithm. The results confirmed that, on average, diagnostic accuracy was improved and diagnosis time was shortened in all proficiency groups significantly.

In the second study, a surgical instruments tracking algorithm for robotic surgery video was developed, and a model for quantitatively evaluating the surgeon’s surgical skill based on the acquired motion information of the surgical instruments was proposed. The movement of surgical instruments is the main component of evaluation for surgical skill. Therefore, the focus of this study was develop an automatic surgical instruments tracking algorithm, and to overcome the limitations presented by previous methods. The instance segmentation framework was developed to solve the instrument occlusion issue, and a tracking framework composed of a tracker and a re-identification algorithm was developed to maintain the type of surgical instruments being tracked in the video. In addition, algorithms for detecting the tip position of instruments and arm-indicator were developed to acquire the movement of devices specialized for the robotic surgery video. The performance of the proposed method was evaluated by measuring the difference between the predicted tip position and the ground truth position of the instruments using root mean square error, area under the curve, and Pearson’s correlation analysis. Furthermore, motion metrics were calculated from the movement of surgical instruments, and a machine learning-based robotic surgical skill evaluation model was developed based on these metrics. These models were used to evaluate clinicians, and results were similar in the developed evaluation models, the Objective Structured Assessment of Technical Skill (OSATS), and the Global Evaluative Assessment of Robotic Surgery (GEARS) evaluation methods.

In this study, deep learning technology was applied to colorectal polyp images for a polyp classification, and to robotic surgery videos for surgical instruments tracking. The improvement in clinical performance with the aid of these methods were evaluated and verified.

Keywords: Deep learning, Convolutional neural network, Colonoscopic diagnosis, Surgical instrument tracking, Robotic surgical skill assessment

Student Number: 2015-30264

Contents

Abstract i

Contents iv

List of Tables vii

List of Figures viii

Chapter 1 General Introduction 1

1.1 Deep Learning for Medical Image Analysis 1

1.2 Deep Learning for Colonoscipic Diagnosis 2

1.3 Deep Learning for Robotic Surgical Skill Assessment 3

1.4 Thesis Objectives 5

Chapter 2 7

Optical Diagnosis of Colorectal Polyps using Deep Learning with Visual Explanations

2.1 Introduction 7

2.1.1 Background 7

2.1.2 Needs 8

2.1.3 Related Work 9

2.2 Methods 11

2.2.1 Study Design 11

2.2.2 Dataset 14

2.2.3 Preprocessing 17

2.2.4 Convolutional Neural Networks (CNN) 21

2.2.4.1 Standard CNN 21

2.2.4.2 Search for CNN Architecture 22
2.2.4.3 Searched CNN Training 23
2.2.4.4 Visual Explanation 24
2.2.5 Evaluation of CNN and Endoscopist Performances 25
2.3 Experiments and Results 27
2.3.1 CNN Performance 27
2.3.2 Results of Visual Explanation 31
2.3.3 Endoscopist with CNN Performance 33
2.4 Discussion 45
2.4.1 Research Significance 45
2.4.2 Limitations 47
2.5 Conclusion 49

Chapter 3 50
Surgical Skill Assessment during Robotic Surgery by Deep Learning-based Surgical Instrument Tracking
3.1 Introduction 50
3.1.1 Background 50
3.1.2 Needs 51
3.1.3 Related Work 52
3.2 Methods 56
3.2.1 Study Design 56
3.2.2 Dataset 59
3.2.3 Instance Segmentation Framework 63
3.2.4 Tracking Framework 66
3.2.4.1 Tracker 66
3.2.4.2 Re-identification 68
3.2.5 Surgical Instrument Tip Detection 69

3.2.6 Arm-Indicator Recognition ... 71
3.2.7 Surgical Skill Prediction Model ... 71
3.3 Experiments and Results ... 78
3.3.1 Performance of Instance Segmentation Framework ... 78
3.3.2 Performance of Tracking Framework ... 82
3.3.3 Evaluation of Surgical Instruments Trajectory ... 83
3.3.4 Evaluation of Surgical Skill Prediction Model ... 86
3.4 Discussion ... 90
3.4.1 Research Significance ... 90
3.4.2 Limitations ... 92
3.5 Conclusion ... 96

Chapter 4 Summary and Future Works ... 97
4.1 Thesis Summary ... 97
4.2 Limitations and Future Works ... 98

Bibliography ... 100
Abstract in Korean ... 116
Acknowledgement ... 119

List of Tables

Table 2.1 Baseline characterisitcs of partifipating endoscpists............ 13

Table 2.2 Polyp characteristics of training set 16

Table 2.3 Patient information and polyp characteristics in the test set . 17

Table 2.4 The CNN performance comparison between a previous method and the proposed methods 30

Table 2.5 Diagnostic accuracy stratified based on the viewing condition (non-assisted versus AI-Assisted) 34-35

Table 2.6 Comparison of the diagnostic accuracy according to endoscopic experiences in non-assisted and AI-assisted cases . 36

Table 2.7 Sensitivity and specificity values stratified based on the viewing condition (non-assisted versus AI-assisted) 37-38

Table 2.8 Comparison of average diagnostic times for each polyp image between CNN and endoscopists.. 42

Table 2.9 Mean score for grit (5-scale) and strength of correlation between optical diagnostic accuracy (r = correlation coefficient). 44

Table 3.1 Training datasets for the instance segmentation framework and spatial-temporal re-identification. 61

Table 3.2 Description of test datasets...62-63

Table 3.3 Description of object structured assessment of technical skills (OSATS) with relevance to motion metrics 73-74

Table 3.4 Description of global evaluative assessment of robotic surgery (GEARS) with relevance to motion metrics. 75-76

Table 3.5 Description of motion metrics.. 79

Table 3.6 Comparative performance of re-identification methods 83

Table 3.7 Comparative performance of surgical instruments tip detection 88

List of Figures

Figure 1.1 Convolutional neural network architecture 2

Figure 2.1 Examples of colorectal polyp 8

Figure 2.2 Examples of narrow band image (NBI) of colon polyp 9

Figure 2.3 Study design 12

Figure 2.4 Illustration of experimental condition and polyp samples 15

Figure 2.5 Data acquisition program 18

Figure 2.6 Applied augmentation techniques 20

Figure 2.7 Loss graph of training and validation set 28

Figure 2.8 CNN architecture for the classification of NBI polyps searched based on the method of neural architecture search 29

Figure 2.9 Result of t-SNE to NBI polyp images 31

Figure 2.10 Results of probabilistic diagnosis as a heatmap on polyp images using Grad-CAM 32

Figure 2.11 Improved accuracy of optical diagnosis with AI assistance classified by group 39

Figure 2.12 Comparisons of the diagnostic accuracy outcomes according to endoscopic experiences in non-assisted and AI assisted conditions 40

Figure 2.13 Scatterplots of optical diagnosis with AI assistance 41

Figure 2.14 Scatterplot for AI-assisted optical diagnosis against grit score 43

Figure 3.1 Surgical robot system and training 51

Figure 3.2 Exercises with the dV-Trainer 53

Figure 3.3 Mechanical structural of da Vinci surgical arm 54

Figure 3.4 Examples of complex robotic surgery view 55

Figure 3.5 Overview of the surgical skill assessment system in robotic surgery 58

Figure 3.6 The overview of the instance segmentation framework 65

Figure 3.7 Concept of tracking: Association of objects over frames...... 66
Figure 3.8 The overview of the tracking framework.............................. 67
Figure 3.9 Block-diagram of Deep SORT 68
Figure 3.10 Procedure of the tips of surgical instruments detection 70
Figure 3.11 Qualitative results of the instance segmentation framework 80
Figure 3.12 Plots of model loss on the training and validation datasets .. 81
Figure 3.13 Trajectory of multi-surgical instruments tip.......................... 85
Figure 3.14 Performance evaluation method between algorithm and ground truth 86
Figure 3.15 Comparison of the performance of surgical skill prediction models and parts of items in OSATS and GEARS with a confusion matrix... 87
Figure 3.16 Relative importance of motion metrics in surgical skill prediction model.. 89
Figure 3.17 Errors in the instance segmentation framework.................... 94
Figure 3.18 Errors in the tracking framework 95

Chapter 1 General Introduction

1.1 Deep Learning for Medical Image Analysis

Deep learning is a subset of machine learning methods, and it performs better in various fields than alternative methods due to the utilization of big data, high computing power and advanced algorithms [1]. Improved performance of deep learning was first shown with a deep neural network (DNN) which is an advanced architecture of a traditional artificial neural network (ANN) composed of multiple stacked hidden layers which are better suited to extract features from high dimensional data.

Convolutional neural network (CNN) is a deep learning method widely used for a pattern recognition in images [2]. CNN excels at extracting features from large image datasets such as ImageNet, and has the advantage of using fewer parameters than DNNs. Architecturally, CNN consists of multiple convolution filters and activation functions that work to extract key features from the input data [3]. Based on the extracted features, CNN provide a final result of task, such as classification through the fully connected layers as the output (Figure 1.1).

In this regard, CNN has been widely used to analyze medical images [4-6]. Two representative applications of CNN to medical images include classification of diabetic retinopathy in retinal fundus images [7] and classification of cancer in skin images [8]. Furthermore, CNN was able to diagnostic performance on par to experts in several other medical application areas [4]. CNN also showed improved performance in other tasks, such as detection, segmentation and registration [4]. Furthermore, CNN was also used for generation in generative adversarial network (GAN) which showed promising results in various applications [9].

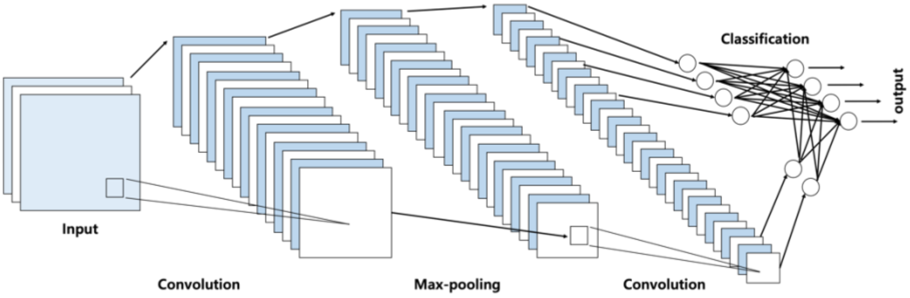


Figure 1.1 Convolutional neural network architecture.

1.2 Deep Learning for Colonoscopic Diagnosis

The application of CNN to various endoscopic images , such as esophagogastroduodenoscopy [3], colonoscopy [10-12], and wireless capsule endoscopy [13], have been reported. For colonoscopy, CNN-based diagnostic methods have been applied to colorectal polyps. Representatively, there are classification algorithms for discriminating the type of colon polyps based on CNN which were as accurate as medical experts [11, 12, 14]. Studies to detect the location of colon polyps [14, 15], to segment the specific area of the polyps

[10, 16] in colonoscopic view, and for real-time inspection [11, 17] were also actively researched. Other related studies that generate high-quality polyp images of imbalanced types for equalizing datasets using GAN have been reported [18].

By applying the CNN to the colonoscopic images, it can help improve the diagnostic accuracy of specific tasks and reduce the diagnosis time, as well as assist in the clinical workflow by performing quantitative image analysis and automatic summary report [19]. Deep learning will have a significant impact on the performance of clinicians using colonoscopy. On the other hand, it is expected that these technological innovations will gradually melt into the medical ecosystem in a way that does not completely replace the role of the clinicians but assist in repetitive and labor-intensive tasks, training students and tasks requiring experience [20].

1.3 Deep Learning for Robotic Surgical Skill Assessment

In the field of surgery, there have been attempts to analyze laparoscopic and robotic surgery videos using CNN [21-24]. These two types of surgical views are similar in that they use surgical instruments in a narrow field of view, and that it is important to recognize the movement of surgical instruments in the view for enhanced situational awareness [25].

One of the methods for acquiring the movement of the surgical instruments is the kinematics approach, which has been used to calculate the relationship between each joint of the robotic arm [26-29], and studies using CNN to analyze the kinematic data have been reported [26].

Another method is the vision-based approach, which has the advantage of recognizing the position of surgical instruments on the screen directly, and increased recognition accuracy using deep learning techniques have been reported in previous studies including classification of the types of instruments [30, 31], detecting the position on the laparoscopic view in real time [22, 32], segmenting specific areas [33], and recognizing joint units using a pose estimation method [24, 34]. Furthermore, by using videos of laparoscopic and robotic surgery, several studies reported on surgical phase identification [21, 35], and surgical action recognition [36].

These studies show that the movement of the surgical instruments represent significant information on each task, and therefore, it may be used as an index to evaluate the surgeon’s surgical skills [37, 38]. If it is possible to automatically obtain the motion of the surgical instruments from the video, the surgeon does not have review the entire surgery process. Furthermore, if a system is developed for quantitatively evaluating the surgical skills based on the movement of a surgical instruments, it will be possible to replace the current subjective evaluation methods based on questionnaires [37, 38], and ultimately help to improve the field as a whole.

1.4 Thesis Objectives

The objective of this study is to improve optical diagnosis using colonoscopic images and surgical skill using robotic surgery videos. In addition, the clinicians who participated in each of task were divided into three groups based on clinical proficiency, and these groups were evaluated to verify the effectiveness of the proposed methods. The improvement in the performance in this study indicates increased optical diagnostic accuracy and the reduced diagnosis time. It may also indicate improvement in robotic surgical skill, and a quantitative evaluation system was developed as the foundation for this purpose.

In chapter 2, we developed a deep learning method that automatically classifies polyp types in colonoscopic images and verified clinical effectiveness. In previous studies, the performance of CNN shows the level of diagnostic accuracy equal to endoscopists. However, even if artificial intelligence (AI) has high accuracy, the final diagnosis is determined by an endoscopist, thus, clinically, AI must be used as an assistive tool and its usefulness must be proved. In this study, a polyp classification algorithm was developed using an AutoML[1] resulting in improved performance over previously reported CNN based methods. In addition, improvements in diagnostic accuracy and diagnosis time

[1] AutoML, automated machine learning

were evaluated for each proficiency group when the endoscopists were assisted by AI.

In chapter 3, we developed an algorithm that automatically tracks the position of surgical instruments in robotic surgery video, and by acquiring motion information of the instruments, we proposed method for quantitatively evaluating a surgeon’s robotic surgery skill. In previous studies, vision-based instrument tracking algorithm has two main limitations. The first deals with the issue of occlusion, specially between surgical instruments, and the other is the issue of maintaining the identity of instruments that change over time. In this study, the two issues were solved using deep learning-based methods, which includes instance segmentation framework and tracking framework. In addition, the tracking accuracy was increased through the arm-indicator recognition algorithm which takes the environment of robotic surgery into account. Finally, a proposed a novel proficiency evaluation model based on the movement of surgical instruments is proposed to replace the existing questionnaire format-based surgical skill evaluation.

Chapter 2

Optical Diagnosis of Colorectal Polyps using Deep Learning with Visual Explanations

2.1 Introduction

2.1.1 Background

Colorectal cancer (CRC) is reported to be the third leading cause of death in the United States [39]. Furthermore, over the past several decades, the incidence of CRC has significantly increased in Asia countries, including Korea [40]. Most CRCs usually develop from preexisting adenomas, which are precancerous lesions, through the adenoma–carcinoma sequence [41]. In this regard, colonoscopy is currently the most important screening test for CRC removal of precancerous adenomatous polyps [42]. This is why adenoma detection is considered a key quality indicator of colonoscopy. Accordingly, considerable research efforts are directed toward the increase of the adenoma detection rate based on physician training and technical advances.

Although the detection and removal of adenoma contribute toward the reduction of CRC, the increased medical costs, including pathological analyses, also must be considered [42]. Most adenomatous polyps detected during screening colonoscopy are diminutive polyps (≤ 5 mm in size). These rarely progress to CRC, however, the current practice is to subject all polyps to pathological evaluation [43, 44]. Diminutive hyperplastic polyps of the rectosigmoid colon are very common benign lesions and do not require removal [45]. Moreover, discrepancies between endoscopic and pathologic diagnoses are not uncommon, and pathological diagnosis is not the gold standard for diagnosing colorectal polyps (≤ 3 mm) [46, 47].

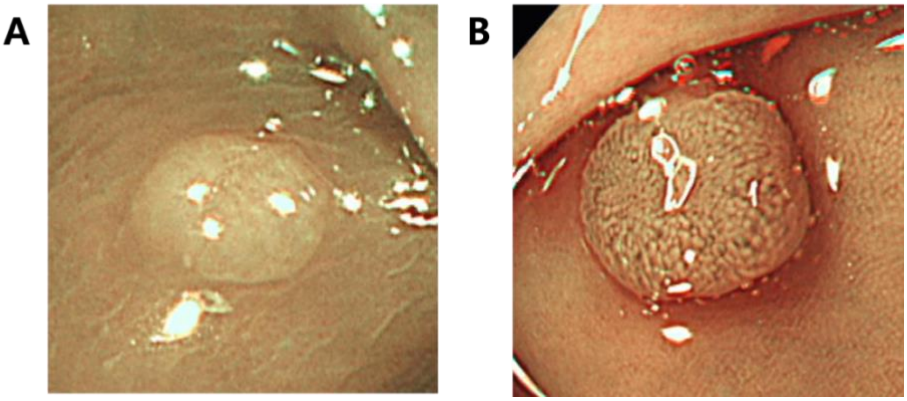


Figure 2.1 Examples of colorectal polyp. (A) Adenoma polyp (B) Hyperplastic polyp

2.1.2 Needs

The application of an accurate endoscopic diagnosis before resection is advantageous because it prevents unnecessary resection and pathological evaluation. In this regard, optical diagnosis based on narrow-band imaging (NBI) can be used to predict the pathology of colorectal polyps and assist the distinction between adenomatous and hyperplastic colorectal polyps [48, 49].

However, this implies that the endoscopist is required to be sufficiently trained to perform adequate optical diagnosis [50]. Furthermore, such optical diagnosis is dependent on the endoscopist's skill and experience [51]. However, this limitation can be overcome with the newly developed computer-aided diagnosis (CADx) [52].

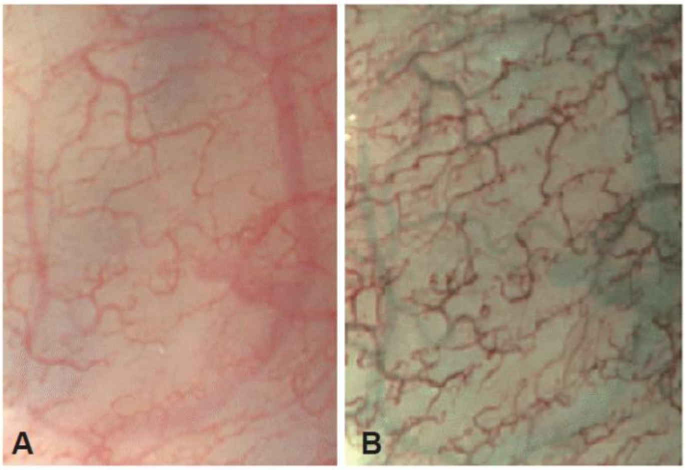


Figure 2.2 Examples of narrow band image (NBI) of colon polyp[2]. (A) Non-NBIs of colon polyp (B) NBIs of colon polyp

2.1.3 Related Work

Recent advances in AI technology have accelerated the development of CADx [20, 53] toward distinction between adenomatous and hyperplastic colorectal polyps [54]. The recent CADx system have demonstrated satisfactory diagnostic capability in predicting the histology based on images

[2] Adapted from "Narrow Band Imaging: Technology Basis and Research and Development History", by Kazuhiro Gono, 2015, Clinical endoscopy, 48, 476. Copyright 2015 by "Korean Society of Gastrointestinal Endoscopy". Adapted with permission.

captured with a magnification endoscope (X80) and an endocytoscope (×500) [54, 55]. The system have reported on classification systems that have demonstrated expert-endoscopist-level accuracy of optical diagnosis [54, 56]. However, these advanced imaging modalities are not commonly used in clinical practice, and the effects of these models applied to endoscopists are not well understood.

Following recent advances in convolutional neural networks (CNN), one of the deep learning approaches, have enabled their use in analyzing medical images [4, 17, 20, 53, 57]. In this regard, many studies have reported on the convergence of the physician's skill and the use of artificial intelligence (AI) to afford accurate diagnoses [20, 58]. In particular, in the optical diagnosis of colorectal polyps, CNN can afford high-performance diagnosis and detection from various colorectal-polyp images [11, 54, 59, 60]. However, even if AI approach affords high-performance colorectal-polyp diagnosis, endoscopists are currently required to perform a final diagnosis for the reasons of safety and accountability, and therefore, it is necessary to verify whether AI-based assistance can effectively aid in the final diagnosis [20]. Recently, Shahidi et al [47] introduced a real-time AI clinical decision support solution and showed that it could help the final diagnoses in the cases in which there were discrepancies between the endoscopic and pathologic diagnoses for diminutive polyps (≤ 3mm).

Our study is the first attempt to identify how the diagnostic capabilities of endoscopists differ between AI-unassisted (test 1) and AI-assisted diagnoses (test 2). Our study showed that AI assistance augments the physician's judgment, thereby improving the accuracy of optical diagnosis and the shortening the diagnostic time. The purpose of this study was to develop a CNN model to determine the pathologic classification of diminutive colorectal polyps on colonoscopic NBIs, and to validate its performance compared to classifications determined by clinical endoscopists. Based on performance comparisons, we investigated the effect of assistive AI technology on the diagnostic accuracy, and compared it with the accuracy associated with the proficiency of endoscopists based on the classification of the polyps in NBIs as either adenomatous or hyperplastic.

2.2 Methods

2.2.1 Study Design

This study was based on a multicenter study conducted from October 2015 to July 2019. It consisted of 3 stages: (1) developed CNN for optical diagnosis of diminutive colorectal polyps, (2) conducted an endoscopic performance assessment and comparison with CNN (test 1), and (3) performed an endoscopic performance assessment with knowledge of the CNN-processed results (test 2).

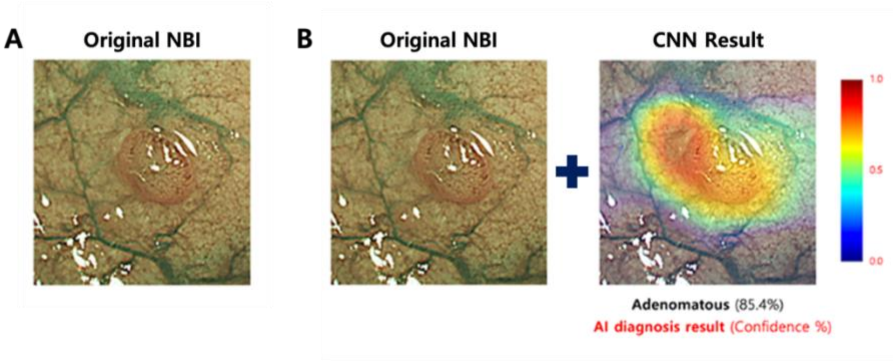


Figure 2.3 Study design. (A) Test 1: Endoscopic performance assessment and comparison with AI (B) Test 2: Endoscopic performance assessment with AI assistance

Twenty two endoscopists participated in this study in 3 groups: (1) novices: 7 gastroenterology trainees with less than 2 years of colonoscopic experience from the Seoul National University Hospital, (2) experts: 4 board-certificated gastroenterologists with various experiences in NBI, and (3) NBI-trained experts: 11 board-certificated gastroenterologists who were trained in optical diagnosis using NBIs, commonly referred to as the Gangnam-READI program [61] (Table 2.1).

The study protocol adhered to the ethical guidelines of the 1975 Declaration of Helsinki and its subsequent revisions, and was approved by the institutional review board (IRB, number H-1702-139-834). Written informed consent was obtained from all participating physicians.

Table 2.1 Baseline characteristics of participating endoscopists.

	n (%)
Sex	
Male	4 (18.2)
Female	18 (81.8)
Colonoscopy experience (years)	
< 2 years	7 (31.8)
2–9 years	6 (27.3)
10–14 years	5 (22.7)
≥ 15 years	4 (18.2)
Estimated cumulative colonoscopy volume	
< 1,000	4 (18.2)
1,000–2,500	4 (18.2)
2,500–4999	5 (22.7)
5,000–9,999	6 (27.3)
≥ 10,000	3 (13.6)
Observed polyp with NBI mode in usual practice	
Not at all	1 (4.5)
> 25 %	4 (18.2)
> 50 %	5 (22.7)
> 75 %	6 (27.3)
All	6 (27.3)
Usefulness of NBI mode for optical diagnosis	
Not at all	0 (0.0)
> 25 %	3 (13.6)
> 50 %	7 (31.8)
> 75 %	7 (31.8)
All	5 (22.7)

2.2.2 Dataset

For the development of CNN for optical diagnosis, we retrospectively collected colonoscopic NBI of diminutive (≤5 mm) polyps from October 2015 to October 2017 at the Seoul National University Hospital, Healthcare System Gangnam. We used the routine pathology report to provide patient care. All polyps were removed using standard techniques and were subsequently evaluated by 1 of the 16 board-certified pathologists at the Seoul National University Hospital. We used image sets that were collected as part of the Gangnam-Real Time Optical Diagnosis (READI) program [61]. All colonoscopies were performed using high-definition colonoscopy (CF-HQ290, Olympus Co, Ltd., Tokyo, Japan) and acquired NBI with or without near-focus magnification. An expert endoscopist reviewed and selected well-focused, high-quality images with appropriate brightness values. If the optical diagnosis of a polyp was not compatible with the histological reports, the images were excluded. Finally, we trained the CNN with a total 1,100 adenomatous polyps and 1,050 hyperplastic polyps from 1,379 patients (Table 2.1). For the test dataset, we prospectively collected 300 polyp images (180 adenomatous polyps and 120 hyperplastic polyps) from January 2018 to May 2019 (Table 2.2). Figure 2.4 shows the polyp samples presented in tests 1 and 2. All 300 NBI polyp images were de-identified and randomly ordered in each test (Table 2.3). The training, validation, and test sets of endoscopic images of NBI polyps exhibited no overlap.

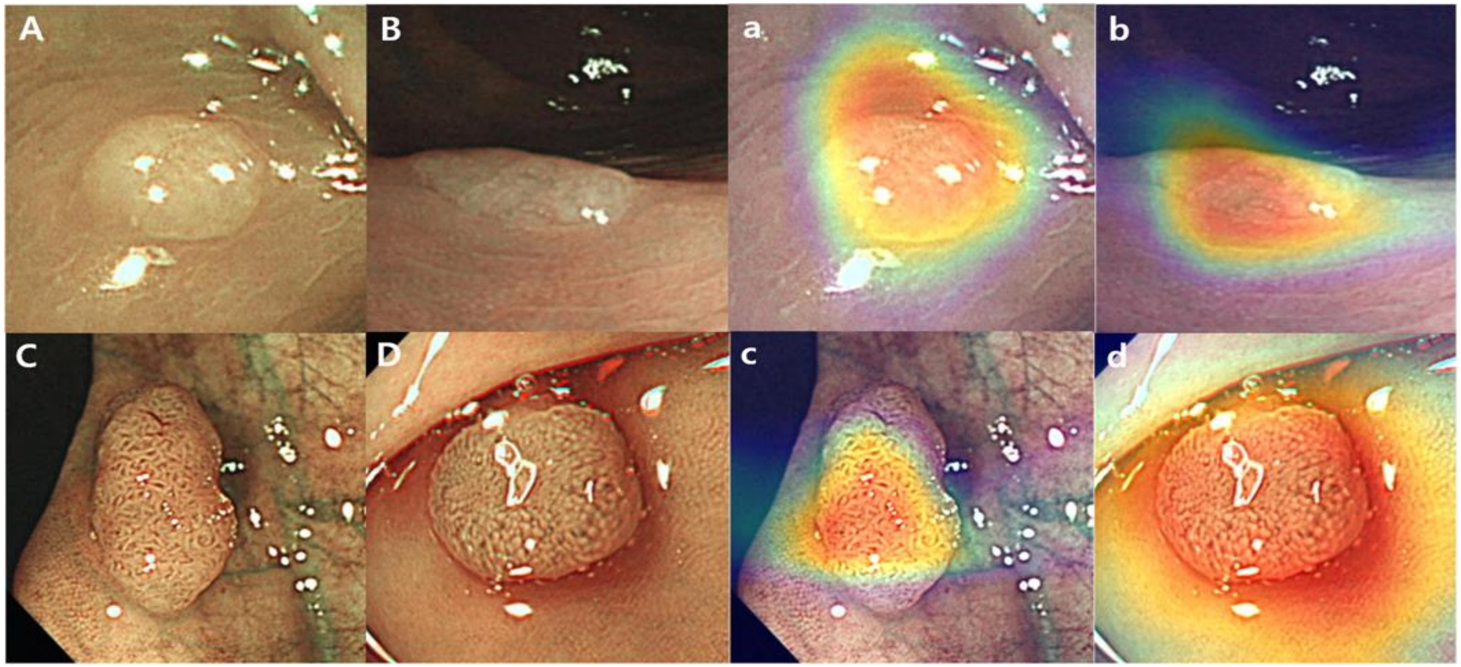


Figure 2.4 Illustration of experimental condition and polyp samples: (A, B, C, D) original NBI images, (a, b, c, d) visual explanation heatmap overlaid on original NBI image. In test 1, we presented the original NBI images, while original NBI images and visual explanation heatmap were presented in test 2.

Table 2.2 Polyp characteristics of training set.

	Adenomatous polyp (N = 1100)	Hyperplastic polyp (N = 1050)	*P*
Location			< 0.0001
- Ascending colon	362 (32.9%)	179 (17.0%)	
- Transverse colon	310 (28.2%)	171 (16.3%)	
- Descending colon	119 (10.8%)	55 (5.2%)	
- Rectosigmoid colon	309 (28.1%)	645 (61.4%)	
Using NF[3] view			< 0.0001
- without NF view	96 (8.7%)	171 (16.3%)	
- with NF view	1004 (91.3%)	879 (83.7%)	
Gross			< 0.0001
- IIa	499 (45.4%)	894 (85.1%)	
- Is	505 (45.9%)	152 (14.5%)	
- Isp	96 (8.7%)	4 (0.4%)	

[3] NF, near focus

Table 2.3 Patient information and polyp characteristics in the test set.

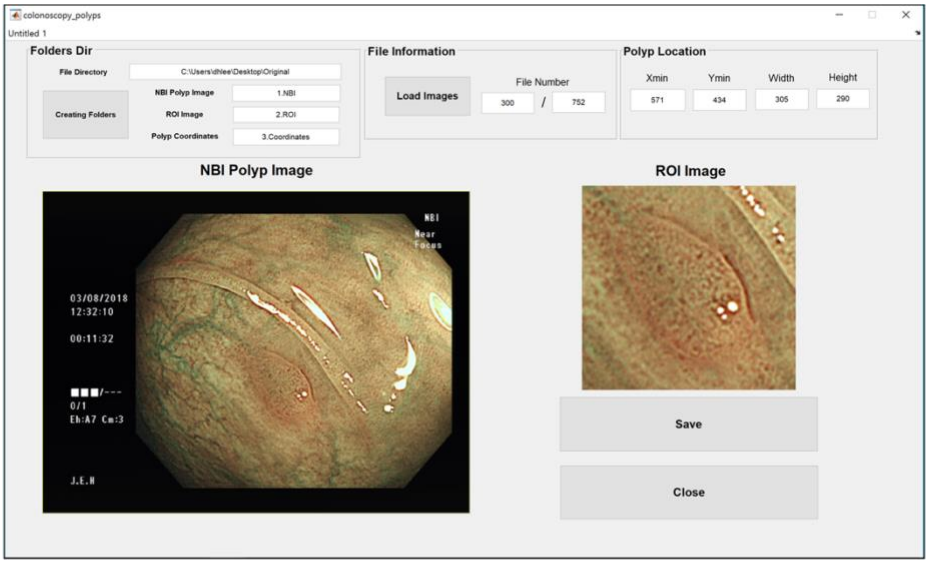
	Adenomatous polyp (N = 180)	Hyperplastic polyp (N = 120)	*P*
Sex			0.062
- Male	127 (70.6%)	97 (80.8%)	
- Female	53 (29.4%)	23 (19.2%)	
Age (mean ± SD)	60.0 ± 10.0	54.9 ± 9.9	0.000
Location			0.000
- Ascending colon	61 (33.9%)	26 (21.7%)	
- Transverse colon	61 (33.9%)	15 (12.5%)	
- Descending colon	14 (7.8%)	13 (10.8%)	
- Rectosigmoid colon	44 (24.4%)	66 (55.0%)	
Using near-focus (NF[4]) view			0.752
- without NF view	12 (6.7%)	10 (8.3%)	
- with NF view	168 (93.3%)	110 (91.7%)	
Gross			0.002
- IIa (flat)	131 (72.8%)	106 (88.3%)	
- Is (sessile)	34 (18.9%)	13 (10.8%)	
- Isp (subpedunculated)	15 (8.3%)	1 (0.8%)	

2.2.3 Preprocessing

The polyp regions-of-interest (ROI) in the images were used for the training, validation, and test were conducted with the developed data acquisition

[4] NF, near focus

program. The program was developed for region-of-interest (ROI) analyses of polyp image acquisitions from original polyp images. This program provides various functionalities, including the ability to import images in a folder, draw ROIs with the mouse, and the save the coordinates of the polyp in the image. The program was developed in MATLAB (MATLAB R2017a, MathWorks Inc., Natick, MA, USA), as shown in Figure 2.5. In the data acquisition step, the NBI has a size of 1280 x 960 pixels (200%), and the ROI of polyp region is cropped within a selected ROI.



colonoscopy_polyps

Untitled 1

Folders Dir

File Directory: C:\Users\dhlee\Desktop\Original

NBI Polyp Image: 1.NBI

ROI Image: 2.ROI

Polyp Coordinates: 3.Coordinates

File Information

File Number: 300 / 752

Polyp Location

Xmin	Ymin	Width	Height
571	434	305	290

Figure 2.5 Data acquisition program. The program provides functionalities for loading original polyp NBIs, selection, and saving.

The shape of the polyp ROI image was square and was resized to 128 × 128 pixels to fit the input size of the CNN. A 5-fold cross-validation was applied as the training step, and the augmentation techniques were applied to generate the training datasets.

As part of the training the convolutional neutral networks (CNN), the augmentation technique is used to improve performance. In this experiment, the number of training sets was increased 5 times based on the application of the augmentation techniques, and yielded the highest performance based on several experiments. The applied methods were a combination of linear transformations (zoom; 0.15, shear; 0.3, rotation; 60 °) and an elastic transformation [62] (σ; 12, random 3×3 gird) using the software packages OpenCV (version 3.4.1) and elasticdeform (version 0.4.6). The results of the augmentation techniques are shown in Figure 2.6.

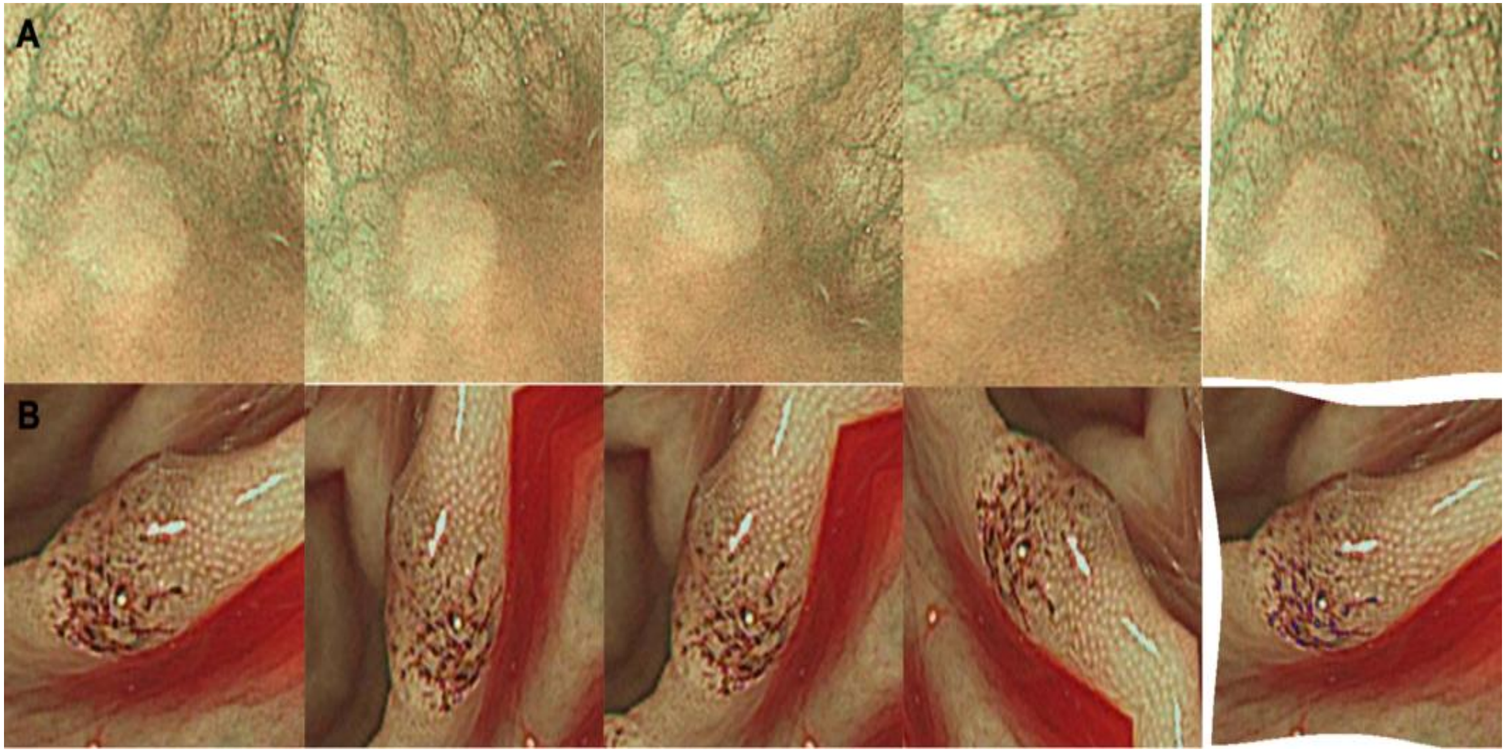


Figure 2.6 Applied augmentation techniques. (A) Augmentation results of hyperplastic polyp images, and (B) augmentation results of adenomatous polyp images.

2.2.4 Convolutional Neural Networks (CNN)

2.2.4.1 Standard CNN

Previous colonoscopic imaging studies using CNN had been selected and trained the defined CNN models, such as inception-v3 [63], which yielded high-performance outcomes in the ImageNet competition. However, these CNN architectures performed tasks on general datasets, and not on specific datasets, such as the NBI polyp. Thus, we used the proposed method to search the CNN architecture by training that was optimized for polyp NBI.

For this reason, the automated machine learning (AutoML) has emerged and overcome the previous limitations and optimized both the network architecture and hyperparameters based on training methods. Generally, AutoML automates machine learning modeling, algorithm selection, and hyperparameter tuning. Selecting and training standard CNN models requires the knowledge and experience of engineers and experimentation based on trials and errors [64]. Therefore, the use of AutoML, represents an attempt to optimize this complex and time-consuming process based on training, commonly referred to as the 'learning to learn' methodology.

2.2.4.2 Search for CNN Architecture

This study used an efficient neural architecture search via parameter sharing (ENAS) [64], which is one of the AutoML methods. ENAS uses recurrent neural networks (RNN) [65] and reinforcement learning (RL) methods [66] to determine the architecture for classifying the specific dataset. In this case, the RNN that determines the architecture of the model is called a controller, and the model created by the controller is called a child network. The Controller used RL method to yield a child network performance based on the accuracy of the generated child network. In turn, the child network trained each sampled child network with a general image training method and with the use of a training dataset.

The proposed method is the architecture searching method and the procedure is as follows. First, the controller RNN generates hyperparameters for the architectural design of CNN. Second, As the controller RNN constructs the architecture, it calculates the accuracy of the validation set based on training until the loss converges. Third, to maximize the expected validation accuracy of the constructed architecture, a policy gradient method which is one of the RL training methods, is used to optimize the hyperparameters of the controller RNN. Finally, This process is repeated to search for the optimal architecture design.

Specifically, this study used a micro search [67] to design small modules and then connected them to CNN. The modules consisted of normal cells and reduction cells, and these 2 modules formed the networks in a repeating architecture.

In addition, 5 types of operations were determined within the modules based on training, and the types were (1) identity, (2) separable convolution with kernel sizes of 3 × 3 and 5 × 5, and (3) average pooling and max pooling with a kernel size of 3 × 3. The hyperparameters for training of the controller RNN and micro search were determined based on experiments as follows. The RNN controller learning rate was 0.003, the child learning rate was 0.0005 to 0.05, the L2 regularization was 1e-4, and the numbers of the child layer, branches, and child cells were 5, 5, and 15, respectively.

2.2.4.3 Searched CNN Training

The training protocol of the model determined by the searching method is as follows. The model was trained with an epoch of 450 with a batch size of 10. An Adam optimizer [68] was used with a learning rate of 0.0001 with decaying using the cosine learning method. In addition, a weighted cross-entropy method [69] was used to solve a class imbalance issue, and the ratio of training datasets was not precisely 1:1.

This study compared the performance between inception-v3 [63], used in a previous study [11, 54], and the proposed method. Furthermore, we compared the results of the ENAS with those of the training set with the use of an augmentation method. The comparison of the performance outcomes include the accuracy, sensitivity, specificity, negative and positive predictive values, and diagnosis time.

The hardware development environment included the NVIDIA Titan V, graphics processing unit, and the software was Python (version 3.4.2; Python Software Foundation, Beaverton, Ore), TensorFlow (version 1.11.0; Google, Mountain View, California, USA). It was developed with reference to https://github.com/melodyguan/enas/.

2.2.4.4 Visual Explanation

The diagnostic confidence (probability) of hyperplastic and adenomatous polyps, which are the results of softmax in an inference step, were presented in a prospective study. In addition, a method of gradient-weighted class activation mapping (Grad-CAM) [70] was used to indicate the location of probabilistic evidence, and a heatmap overlaid on the polyp images diagnosed by the CNN was presented in a prospective study. Grad-CAM is one of the explainable AI techniques that presents the results of the CNN as a probabilistic representation of a heatmap overlaid on an image. The closer the color of the heatmap is to blue, the lower is the probability, and the closer the color is to red, the higher is the probability.

Additionally, t-Distributed stochastic neighbor embedding (t-SNE) [71] is a dimension reduction method, whereby high-dimensional data is embedded as low-dimensional data and are visualized. We defined the similarity between the data in a high-dimensional space represented by probability values and the similarity between the data in an embedding (low-dimensional) space. Accordingly, the gradient descent was used so that the difference between the 2 similarities is small.

In this study, features of the validation set were extracted from the last layer of the trained CNN. The number of features was 1024, and the features of the last layer-1 reduced the dimension to 2, with a learning rate of 200, and with 1000 iterations based on the use of the package scikit-learn machine learning package (version 0.19.1; https://scikit-learn.org).

2.2.5 Evaluation of CNN and Endoscopist Performances

The following 2-stage tests were conducted based on the use of the validation dataset. In test 1, each endoscopist independently evaluated the digital format of polyp NBIs to determine whether the polyp was adenomatous or hyperplastic test set on a retina display of a computer via on online survey. After a month, they performed test 2 in the same way as the previous test 1. In test 2, each endoscopist made an optical diagnosis based on the original polyp NBI (test 1) and the CNN-processed results. The AI results presented to the physician were as follows: (1) The AI predicted the pathology (adenomatous or hyperplastic polyps), (2) confidence value, and (3) both original NBI polyp and an explanation heatmap of the NBI polyp obtained using Grad-CAM (Figure 2.4). In addition, each test also recorded the start and end times to calculate the average diagnostic time per polyp image.

The optical diagnosis performances of the CNN and endoscopist (test 1), and those of the endoscopists with AI assistance (test 2) were evaluated and compared with the use of the McNemar test. We developed a mixed-effects logistic regression model to estimate the effect of AI assistance on the subgroups. Wilcoxon signed rank tests were used to assess differences of diagnostic time between nonassisted and AI-assisted assessments. For all tests, a *P* value of 0.05 was considered to indicate statistical significance, and a *P* value correction was performed. All calculations were performed using SAS (version 9.3; SAS Institute, Cary, NC) software package.

After 2 tests, we conducted individual surveys for the personality characteristics with the use of Grit-Original (Grit-O, Korean version) with 2 components, namely, consistency of interest and perseverance of effort [72]. Grit is a positive, noncognitive personality trait characterized by the ability to persevere during difficulties combined with powerful motivation to achieve a goal [73]. Grit has been found to be a superior predictor of success in high achievement fields [74]. Higher grit has been found to correlate with higher performance in medical school, whereas lower grit has been found to correlate with increased surgical residency training drop-out rates [75, 76]. Previous studies have shown that doctors exhibit an average grit score in the range of 3.5 to 3.7 [73, 77]. The Grit-Original was validated based on a questionnaire that comprised 12-items. It was scored on a 5-point scale (from 1 to 5). The summed score was divided by 12 to yield the final Grit score [78].

2.3 Experiments and Results

2.3.1 CNN Performance

Figure 2.7 shows the loss graph of the training and the validation datasets. In the process of training, learning rate decay using cosine learning method was applied. The two loss patterns over epoch show similar trends and there are no significant differences between the two which shows that there is minimal overfitting.

In addition, Figure 2.8 represents the results of the searched CNN architecture based on training, which consists of repeating normal cells and reduction cells. It is a structure in which two streams are connected to calculate the loss function, and each normal cell and reduction cell are composed of a combination of identity, separable convolution (with kernel sizes of 3 × 3, 5 × 5), average pooling, and max pooling.

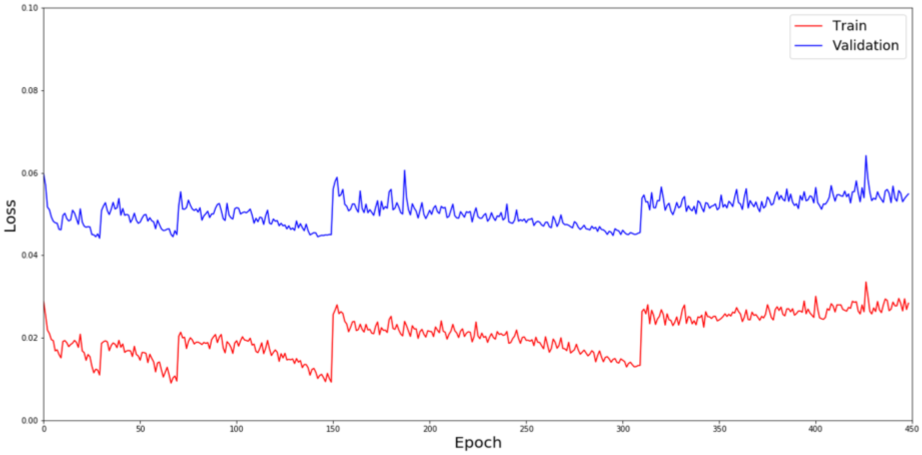


Figure 2.7 Loss graph of training and validation set.

In this study, the CNN selected using ENAS with augmentation techniques exhibited an optical diagnostic accuracy of 86.7% (95% confidence interval 82.3–90.3), with a sensitivity of 83.3% and a specificity of 91.7%. The diagnostic performance of the CNN was compared with those of 22 endoscopists as shown in Table 2.4.

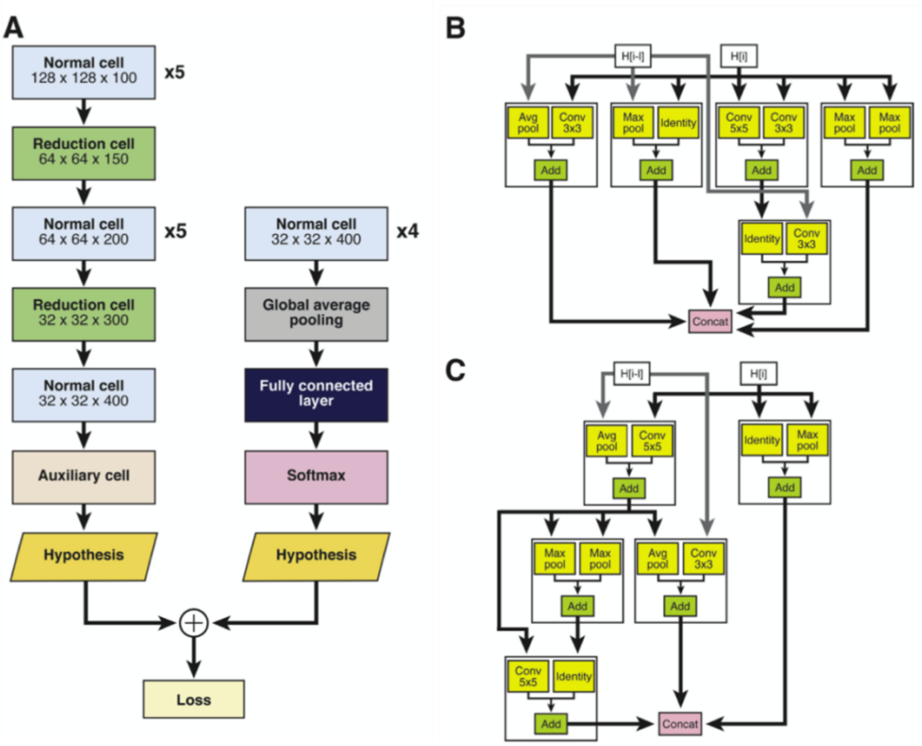


Figure 2.8 CNN architecture for the classification of NBI polyps searched based on the method of neural architecture search. (A) Full architecture of convolutional neural networks searched by the proposed method, (B) architecture of normal cell, (C) architecture of reduction cell

Table 2.4 The CNN performance comparison between a previous method and the proposed methods.

	Accuracy n (%)	Sensitivity n (%)	Specificity n (%)	Positive predictive value n (%)	Negative predictive value n (%)	Diagnostic time (s)
Inception-v3	245/300 (81.67%)	144/180 (80%)	101/120 (84.17%)	141/160 (88.34%)	103/140 (73.72%)	8.42/300
ENAS*	256/300 (85.33%)	147/180 (81.67%)	109/120 (90.83%)	145/160 (90.83%)	107/140 (76.76%)	3.62/300
ENAS* + Augmentation	260/300 (86.7%)	150/180 (83.3%)	110/120 (91.7%)	150/160 (93.8%)	110/140 (78.6%)	

2.3.2 Results of Visual Explanation

Figure 2.9 represents the results of t-Distributed stochastic neighbor embedding (t-SNE), one of the visual explanations methods for the interpretation of CNN. This is a method to visualize the performance of the CNN model, and shows the classification result of embedded high-dimensional features from the 1024 features in the last layer, in two dimensions.

Additionally, Figure 2.10 represents the results of Grad-CAM, one of the visual explanations methods as well. Probabilistic diagnosis shown as a heatmap on polyp images represents the basis for the CNN model prediction.

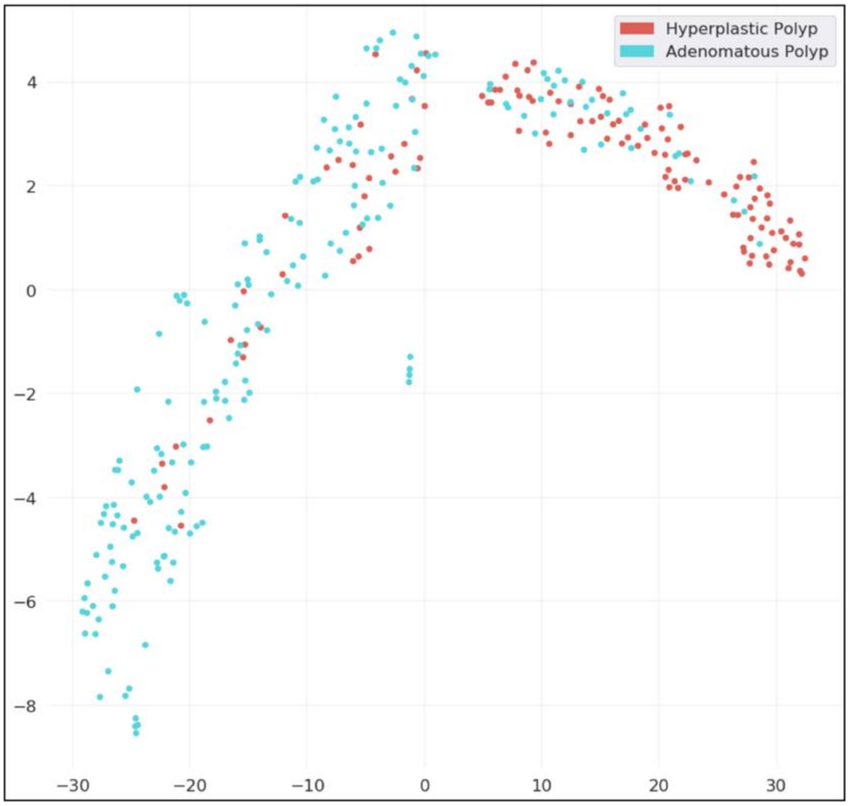


Figure 2.9 Result of t-SNE to NBI polyp images.

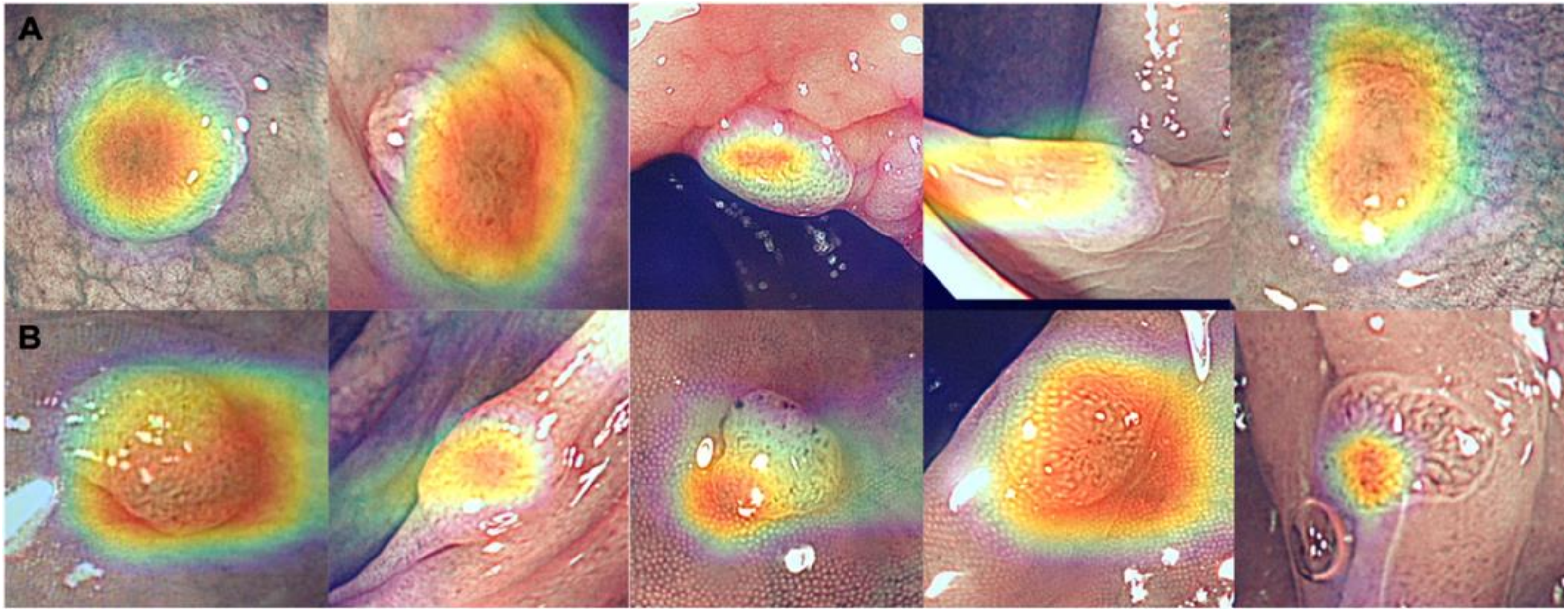


Figure 2.10 Results of probabilistic diagnosis as a heatmap on polyp images using Grad-CAM. (A) Heatmap results overlaid on hyperplastic polyp images. (B) Heatmap results on adenomatous polyp images

2.3.3 Endoscopist with CNN Performance

The diagnostic performance of the CNN was compared with those of 22 endoscopists (Table 2.5). Five of 7 novices yielded significantly lower diagnostic accuracies (47.7–79.0%) than the CNN ($P < 0.05$). Only 1 endoscopist (E1, 77.3%) of the 4 experts demonstrated significantly lower diagnostic accuracy than the CNN ($P < 0.05$). Among the 11 NBI-trained expert endoscopists, 1 endoscopist (N-TE4, 92.7%) demonstrated statistically higher diagnostic accuracy than the CNN ($P = 0.011$).

The overall accuracy of optical diagnosis was significantly increased with the use of AI assistance (82.5% to 88.5%, $P < 0.05$) as shown in Table 2.6. Although the AI assistance appeared to improve endoscopist performance, it must be considered that this increase can vary according to the endoscopist experiences. In the novice group, all endoscopists domenstrated performances with significantly increased accuracies ($P < 0.05$), and 4 of them demonstrated performances with greater accuracy than the algorithm. In the expert group, two endoscopists significantly improved the accuracies (E1, $P = 0.01$; E4, $P = 0.001$), and 1 (E4) achieved higher accuracy than the algorithm. In the NBI-trained expert group, 3 endoscopists (N-TE1, N-TE2, N-TE11) demonstrated performances with significantly improved accuracies ($P < 0.05$). Interestingly, 1 endoscopist (N-TE2) was already more accurate than the algorithm without AI assistance.

Table 2.5 Diagnostic accuracy stratified based on the viewing condition (non-assisted versus AI-Assisted).

	Non-assisted (T1)			T1 versus AI	AI-assisted (T2)			T1 versus T2
	Accuracy			(*P* value)	Accuracy			(*P* value)
Observer	n	Percent	95% CI		n	Percent	95% CI	
AI	260/300	86.7%	(82.3, 90.3)					
Novice (N = 7)								
N1	236/300	78.7%	(73.6, 83.2)	0.009	264/300	88.0%	(83.8, 91.5)	<.0001
N2	237/300	79.0%	(73.9, 83.5)	0.003	261/300	87.0%	(82.7, 90.6)	0.001
N3	245/300	81.7%	(76.8, 85.9)	0.075	262/300	87.3%	(83, 90.9)	0.024
N4	255/300	85.0%	(80.4, 88.8)	0.522	269/300	89.7%	(85.7, 92.9)	0.035
N5	226/300	75.3%	(70.1, 80.1)	<.0001	247/300	82.3%	(77.5, 86.5)	0.007
N6	143/300	47.7%	(41.9, 53.5)	<.0001	237/300	79.0%	(73.9, 83.5)	<.0001
N7	207/300	69.0%	(63.4, 74.2)	<.0001	258/300	86.0%	(81.6, 89.7)	<.0001
Expert endoscopist (N = 4)								
E1	232/300	77.3%	(72.2, 81.9)	0.002	259/300	86.3%	(81.9, 90)	0.001
E2	254/300	84.7%	(80.1, 88.6)	0.460	263/300	87.7%	(83.4, 91.2)	0.208
E3	265/300	88.3%	(84.1, 91.7)	0.515	270/300	90.0%	(86, 93.2)	0.466
E4	254/300	84.7%	(80.1, 88.6)	0.439	276/300	92.0%	(88.3, 94.8)	0.001
NBI-trained expert endoscopist (N = 11)								
N-TE1	258/300	86.0%	(81.6, 89.7)	0.808	276/300	92.0%	(88.3, 94.8)	0.011
N-TE2	264/300	88.0%	(83.8, 91.5)	0.593	282/300	94.0%	(90.7, 96.4)	0.004

N-TE3	267/300	89.0%	(84.9, 92.3)	0.362	265/300	88.3%	(84.1, 91.7)	0.746
N-TE4	278/300	92.7%	(89.1, 95.3)	0.011	282/300	94.0%	(90.7, 96.4)	0.371
N-TE5	269/300	89.7%	(85.7, 92.9)	0.225	271/300	90.3%	(86.4, 93.4)	0.683
N-TE6	264/300	88.0%	(83.8, 91.5)	0.617	268/300	89.3%	(85.3, 92.6)	0.505
N-TE7	270/300	90.0%	(86, 93.2)	0.181	280/300	93.3%	(89.9, 95.9)	0.059
N-TE8	263/300	87.7%	(83.4, 91.2)	0.714	266/300	88.7%	(84.5, 92)	0.602
N-TE9	256/300	85.3%	(80.8, 89.1)	0.537	256/300	85.3%	(80.8, 89.1)	1.000
N-TE10	250/300	83.3%	(78.6, 87.4)	0.211	259/300	86.3%	(81.9, 90)	0.150
N-TE11	252/300	84.0%	(79.4, 88)	0.339	270/300	90.0%	(86, 93.2)	0.011

Table 2.6 Comparison of the diagnostic accuracy according to endoscopic experiences in non-assisted and AI-assisted cases.

	Non-assisted (T1)		AI-assisted (T2)		T1 versus T2				
	Accuracy (%)	SE[5]	Accuracy (%)	SE	Difference	Lower	Upper	SE	*P*
Group									
Novice	73.8	2.86	85.6	1.19	11.86	7.27	16.45	2.19	<.0001
Expert	83.8	3.78	89.0	1.57	5.25	-0.82	11.32	2.90	0.0861
NBI-trained expert	87.6	2.28	90.2	0.95	2.55	-1.11	6.21	1.75	0.1619
Overall	82.5	1.61	88.5	0.67	6.00	3.41	8.59	1.24	0.0001

[5] SE, standard error

Table 2.7 Sensitivity and specificity values stratified based on the viewing condition (non-assisted versus AI-assisted).

	Sensitivity			Specificity		
Observer	Non-assisted (T1)	AI-assisted (T2)	T1 versus T2	Non-assisted (T1)	AI-assisted (T2)	T1 versus T2
AI	83.3%		(*P* value)	91.7%		(*P* value)
Novice (N = 7)						
N1	98.3%	95.0%	0.0578	49.2%	77.5%	<.0001
N2	80.0%	88.3%	0.009	77.5%	85.0%	0.0495
N3	90.6%	92.2%	0.5316	68.3%	80.0%	0.0164
N4	93.9%	90.0%	0.1266	71.7%	89.2%	<.0001
N5	87.8%	85.6%	0.4497	56.7%	77.5%	<.0001
N6	47.8%	80.0%	<.0001	47.5%	77.5%	<.0001
N7	88.3%	85.6%	0.3532	40.0%	86.7%	<.0001
Expert endoscopist (N = 4)						
E1	71.1%	82.8%	0.0027	86.7%	91.7%	0.1336
E2	91.7%	89.4%	0.4328	74.2%	85.0%	0.0093
E3	92.8%	86.7%	0.0343	81.7%	95.0%	0.0003
E4	82.8%	90.0%	0.0158	87.5%	95.0%	0.0126
NBI-trained expert endoscopist (N = 11)						
N-TE1	96.7%	93.3%	0.1088	70.0%	90.0%	<.0001
N-TE2	96.7%	93.3%	0.1088	75.0%	95.0%	<.0001
N-TE3	85.6%	82.2%	0.2568	94.2%	97.5%	0.2059

N-TE4	97.2%	95.6%	0.2568	85.8%	91.7%	0.0522
N-TE5	88.3%	89.4%	0.6547	91.7%	91.7%	1
N-TE6	93.3%	95.0%	0.4054	80.0%	80.8%	0.8348
N-TE7	93.3%	93.9%	0.7815	85.0%	92.5%	0.0201
N-TE8	98.3%	95.0%	0.0578	71.7%	79.2%	0.0606
N-TE9	82.8%	81.7%	0.6831	89.2%	90.8%	0.4142
N-TE10	77.8%	82.2%	0.1441	91.7%	92.5%	0.7389
N-TE11	96.1%	90.0%	0.0076	65.8%	90.0%	<.0001

The optical diagnostic performances of novices, expert endoscopists, and NBI-trained expert endoscopists, were 73.8%, 83.8%, and 87.6%, respectively, and their diagnostic accuracies improved with AI assistance (85.6%, 89.0%, and 90.2%, respectively) as shown in Figure 2.11. Without any AI assistance (test 1), the novice group demonstrated a significantly lower accuracy than both the experts (*P* = 0.049) and the NBI-trained experts (*P* = 0.001). With AI assistance (test 2), the accuracy of the novices significantly improved, and there was no statistical difference when performance were compared with those of the expert group (*P* = 0.102) as shown in Figure 2.12.

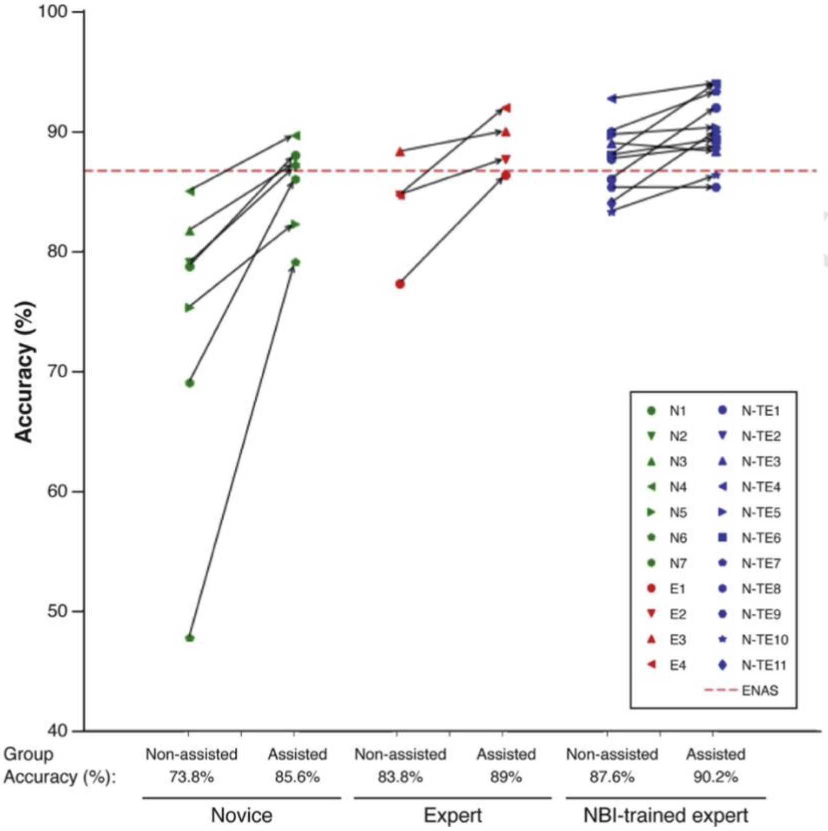


Figure 2.11 Improved accuracy of optical diagnosis with AI assistance classified by group.

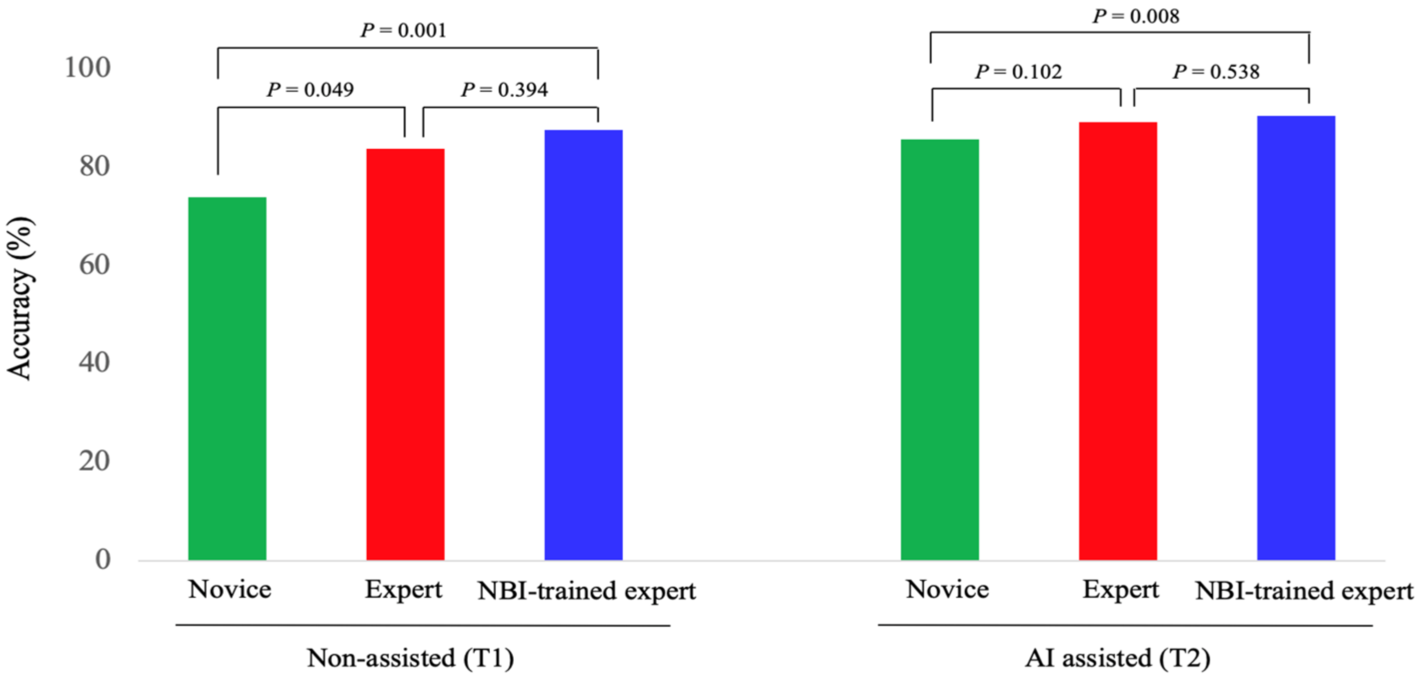


Figure 2.12 Comparisons of the diagnostic accuracy outcomes according to endoscopic experiences in non-assisted and AI assisted conditions.

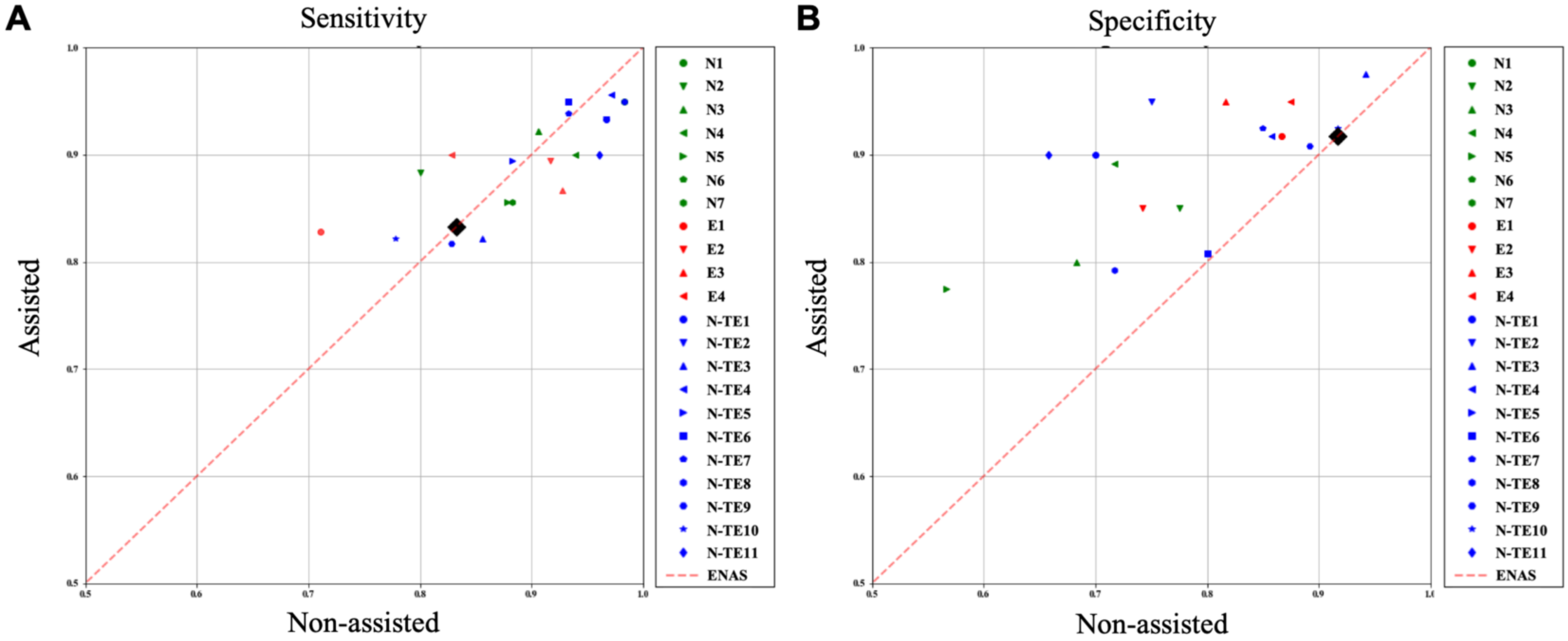


Figure 2.13 Scatterplots of optical diagnosis with AI assistance. (A) sensitivity (B) specificity for each the AI-assisted condition (y-axis) compared with non-assisted condition (x-axis) for participating endoscopists. Results show that AI assistance increased specificity.

The average time for the Al algorithm to diagnose each polyp was 0.01 second, which is significantly shorter than the time taken by the endoscopists (Table 2.8). Herein, we note that AI assistance offered an interpretable explanations such that endoscopists can diagnosis faster. In particular, the diagnostic time per polyp reduced from 4.44 to 3.68 seconds in the case of the NBI-trained expert group (*P* = 0.33).

Table 2.8 Comparison of average diagnostic times for each polyp image between CNN and endoscopists.

	Diagnostic Time per Polyp (s)		
	Non-assisted (T1)	AI-assisted (T2)	*P* value
CNN[6]	0.01	0.01	1.000
Overall	3.92	3.37	0.042
Novice	3.24	3.18	0.866
Expert	3.67	2.84	0.068
NBI-trained Expert	4.44	3.68	0.033

The acceptance of AI assistance by the endoscopist also forms an important factor in diagnosis. This acceptance factor can be reflected by the personality trait of the grit. The traits of the grit are defined as the perseverance and passion for long-term goals, and they reflect the ability of an individual to sustain long-term efforts and overcome obstacles in realizing goals [77]. In our study, the mean participant grit score was 3.56 (Table 2.9). Overall, we observed a

[6] CNN, convolutional neural networks

moderate correlation between grit and AI- assisted diagnostic accuracy (r = 0.51, P = 0.015). Conversely, there was no correlation between grit and diagnostic accuracy without AI assistance.

Our study findings show that endoscopists with high grit scores could flexibly accept AI assistance, thereby increasing the diagnostic accuracy. In this study, we found that high grit, particularly in terms of the consistency of interest, correlated with high accuracy, which translated to a passion to achieve and maintain strong motivation for overcoming obstacles. This result indicates the possibility that certain personality traits of the endoscopist can affect the acceptance of AI technology.

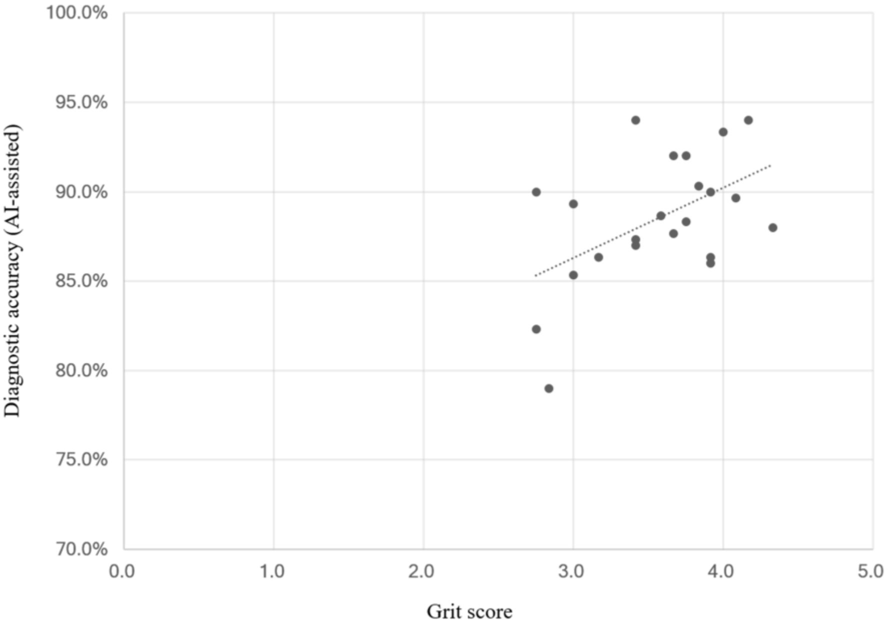


Figure 2.14 Scatterplot for AI-assisted optical diagnosis against grit score.

Table 2.9 Mean score for grit (5-scale) and strength of correlation between optical diagnostic accuracy (r = correlation coefficient).

	Overall			Optical Diagnostic Accuracy			
				Non-assisted (T1)		AI-assisted (T2)	
	Mean	SD[7]	IQR[8]	Correlation, r	*P* value	Correlation, r	*P* value
Grit score	3.561	0.47	3.22-3.92	0.3	0.1768	0.51	0.0148
Consistency of Interest	3.386	0.59	3.00-3.83	0.38	0.0799	0.56	0.0069
Perseverance of effort	3.735	0.5	3.50-4.00	0.11	0.6175	0.31	0.1651

[7] SD, standard deviation

[8] IQR, interquartile range

2.4 Discussion

2.4.1 Research Significance

In this study, we investigated the effect of AI assistance on 22 endoscopists in accurately predicting the pathology of polyp NBI. We found that AI assistance with an interpretable explanations could improve both the optical diagnostic accuracy and the diagnostic speed regardless of the endoscopic experiences. The diagnostic accuracy increased maximally in the novice group, and it was not significantly different from that of the expert group. Herein, we note that AI assistance can aid even NBI-trained expert endoscopists in increasing their diagnostic accuracies and reduce the diagnostic duration.

In a technical point of view, unlike the general method of training standard CNN, we used an ENAS [64], which is one of the AuotML methods, to search the CNN architecture by training that was optimized for polyp NBI. In addition, the proposed method is faster in formulating a diagnosis compared with previous studies given that it is based on a better graphics processing unit performance, smaller batch size, and smaller training image size. Accordingly, it is considered to be suitable for real-time diagnosis. We also found that the diagnostic performances of the ENAS with the augmentation techniques for the flat polyp cases were improved compared to the single-ENAS method in conjunction with the endoscopist diagnoses [62, 79]. Considering that AI did not recognize this type of polyp well in previous studies [80], the use of the proposed methods confirmed that the combination of various augmentation techniques could compensate for the lack of training data and improve the performance.

In the application of medical AI technology, it is important that physicians can understand the AI results to accept AI. Here, we mention that deep learning methods are “black boxes” because it is impossible to explain why the AI arrived at a specific decision [81]. In this context, we note that recently AI explanation methods have been developed to enable humans to comprehend how the AI predictions are made [58, 82]. In this study, we presented the AI results to physicians in the following manner: AI-predicted pathology with confidence value and both original polyp NBI and NBI with generated heatmaps using Grad-CAM methods [70]. We visualized the highlights that overlaid the polyp NBI for predicted evidence. This interpretable explanation of AI results can aid the endoscopists to accept AI assistance, thereby contributing to increase diagnostic accuracy.

On the contrary, in a clinical point of view, we found that AI assistance is most effective in aiding novice rather than experts. All the novice demonstrated significantly increased diagnostic accuracy, and their results were not inferior to those of experts. It should be noted that in previous studies, “nonexpert” results showed marked interobserver variability, and these nonexperts could not achieve acceptable accuracy in the optical diagnosis of diminutive polyps with NBI [83, 84]. To overcome this limitation, many researchers have attempted to develop AI diagnostic algorithms that can allow nonexperts to demonstrate improved accuracy of optical diagnosis as a clinical tool [85-87]. Our study demonstrated that AI assistance may aid in augmenting the abilities of nonexperts with limited training in optical diagnosis to take better decisions.

In our study, the 11 NBI-trained experts participated in the training program for optical diagnosis using NBI from September 2015 to September 2016 [61]. This NBI-trained expert group demonstrated the highest accuracy of 87.6%, which is thought to be attributed to the effects of training. Even in the case of NBI-trained experts whose performance was better than AI algorithms, the accuracy increased and the diagnosis time reduced with AI assistance. These results suggest that AI assistance can also be useful for experts in actual clinical situations.

2.4.2 Limitations

This study has several limitations. First, we developed a CNN based on high-quality images. However, in clinical practice, the acquired images may be of poor quality, such as out-of-focus or blurred images.

Second, this study does not focus on real-time optical diagnoses. We conducted only 2 in vitro tests to compare the performances of endoscopists with and without AI assistance. We cropped and resized images to fit the CNN's input size. These hand-crafted, extracted images could be different from actual colonoscopy images. In actual colonoscopy, the endoscopist could observe polyps at various angles and in continuous frames to predict pathology. Endoscopic video streams could be more useful than still images [11].

Third, our training and test datasets consisted of tubular adenoma with low-grade and hyperplastic polyps. We excluded diminutive polyps with serrated lesions, and other benign conditions, such as inflammatory polyps or lymphoid follicles. Further studies are needed on other types of colorectal polyps with various pathological findings.

Fourth, the confidence value, the probabilistic diagnosis of CNN, is not always reliable because the diagnosis is not based on the same approach as that used for humans [81]. Therefore, to solve the uncertainty issue [86], Bayesian deep learning method has been studied that can be trained with weights of probability distribution rather than with the use of fixed-weight CNN values [88].

Finally, because Grad-CAM is a technology that was applied independently on the proposed CNN architecture and on the training methods, the presented heatmap results were not stable because the results were different at each CNN layer.

2.5 Conclusion

In conclusion, AI assistance is useful for the improvement the accuracy of the optical diagnosis of diminutive polyps and for the achievement of shorter diagnostic times. In particular, we found that AI assistance was most effective for novices because they could achieve accuracies similar to those of experts without training or effort. In this manner, by reducing the diagnostic-capability differences between physicians, pathologic examinations can be replaced by accurate optical diagnoses with AI assistance that can contribute to significant reductions of medical costs.

* Large sections of this chapter were published previously in *Gastroenterology*. (Eun Hyo Jin, Dongheon Lee, et al. “Improved Accuracy in Optical Diagnosis of Colorectal Polyps Using Convolutional Neural Networks with Visual Explanations”. *Gastroenterology*, 2020;158(8):2169-2179) [89]

Chapter 3

Surgical Skill Assessment during Robotic Surgery by Deep Learning-based Surgical Instrument Tracking

3.1 Introduction

3.1.1 Background

The da Vinci system of robotic surgery is a well-established robotic surgical platform that has been deployed worldwide for the past two decades [90]. The da Vinci robot has the advantages of 3-dimentional vision, magnified surgical view, endo-wrist instruments, tremor filtering support, and motion scaling [90, 91]. Robots are therefore performing many types of minimally invasive surgery, including in general, urologic, gynecologic, and cardiothoracic surgery [91].

Most types of robotic surgery require training, with a classic learning curve eventually resulting in consistent performance [92]. Therefore, studies on efficient surgeon training methods according to the learning curve have been reported in robotic surgery [28, 93, 94].

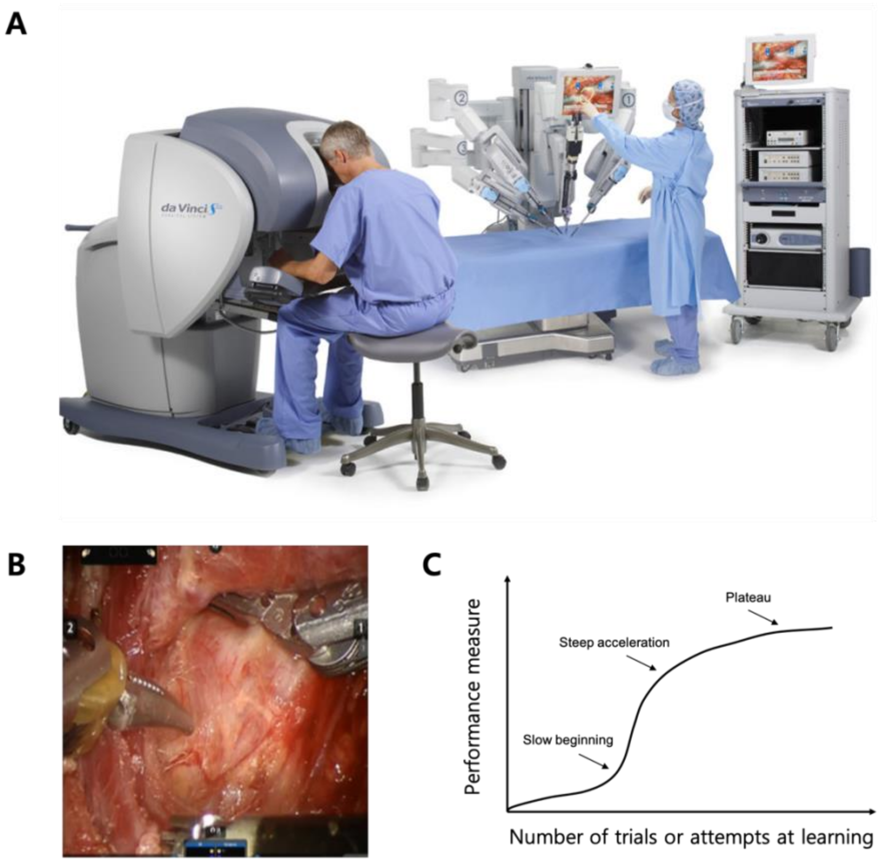


Figure 3.1 Surgical robot system and training. (A) da Vinci surgical system (B) Robotic surgery view (C) Classic learning curve

3.1.2 Needs

It is important to repeatedly evaluate the surgical skill of each surgeon learning robotic surgical procedures to determine surgeon's current position on the learning curve. Surgical proficiency in laparoscopic surgery, a type of minimally invasive surgery, has been evaluated by the Objective Structured

Assessment of Technical Skill (OSATS) developed in 1997 [37]. Moreover, proficiency in robotic surgery, currently the primary type of micro-invasive surgery, has been evaluated by the Global Evaluative Assessment of Robotic Skills (GEARS) [38] developed in 2012.

Qualitative assessment methods such as OSATS and GEARS are subjective, being based on questionnaires [38, 95]. In addition, these methods have limitations in that surgeons need to see and evaluate long-term surgical procedures. An automatic and quantitative method of evaluation for robotic surgery is needed therefore to overcome the limitations of these subjective methods [96]. The main items of OSATS and GEARS are related to the movement of surgical instruments (SIs), resulting in enhanced situational awareness [25]. Application of an SI tracking algorithm to surgical images may automate the evaluation of long-term surgical processes. This evaluation may be quantified by determining the motions of the SI and be quantified by determining the motions of the SI.

3.1.3 Related Work

Current methods of evaluating surgical skill during robotic surgery include the use of the da Vinci Skills Simulator (dVSS). The simulator presents various tasks to surgeons, such as ring and rail, in virtual robotic surgery environments and evaluates surgeons' proficiencies based on its inbuilt evaluation criteria [94, 97]. However, virtual robotic surgery is a practice environment for novice surgeons, differing greatly from actual surgical environments.

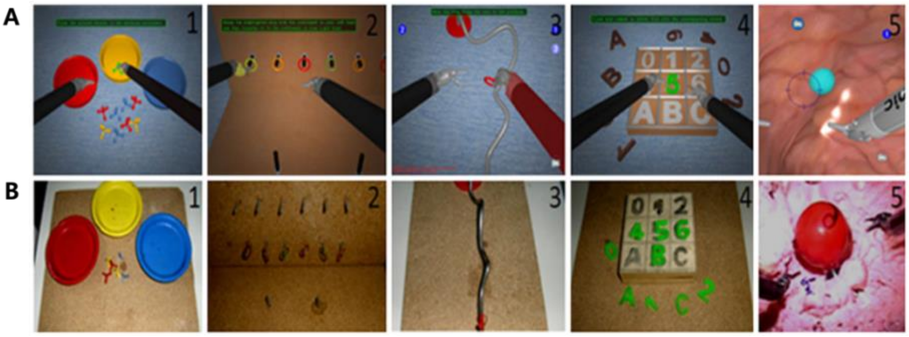


Figure 3.2 Exercises with the dV-Trainer[9]. (A) and on models with the Da Vinci robot (B): Pick and Place, Peg Board, R ing and Rail, Match Board, Camera Targeting

Surgical skill has been determined by quantitatively measuring SI movements in actual surgical environments [27, 28]. Although kinematics methods estimating mechanical movements of SIs have been used to calculate the relationship between each joint of these SIs [26-29] (Figure 3.3), these methods can result in cumulative errors in the calculation of the motions of each joint [25]. Moreover, these methods are inapplicable to most surgical robots, except for some research equipment, because they are prevented from approaching the values of kinematic joints [27, 28, 98].

[9] Adapted from “The virtual reality simulator dV-Trainer® is a valid assessment tool for robotic surgical skills”, by Perrenot, and et al., 2012, Surgical Endoscopy, 26, 2587-2593. Copyright 2012 by “*Springer Nature*”. Adapted with permission.

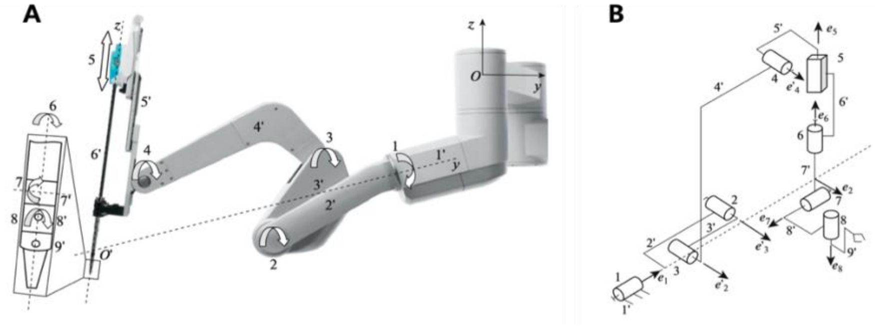


Figure 3.3 Mechanical structural of da Vinci surgical arm[10]. (A) one of the *da Vinci S* manipulators (B) Kinematic scheme of one of the *da Vinci S* manipulators

Image-based methods can directly recognize the SIs in robotic surgery views. Moreover, image-based methods have other advantages because they do not require external equipment and can therefore be applied to surgical robots made by other manufacturers and to laparoscopic surgery.

Traditional image processing approaches, however, are limited in detecting SI tips in complex robotic surgery views [99, 100]. A deep learning-based approach has been found to overcome these limitations and has been applied to several tasks during robotic surgery, such as classification [30, 31], detection [22, 32], segmentation [33], and pose estimation [24, 34] of SIs, phase identification [21, 35], and action recognition [36]. These methods are limited with respect to determining the trajectory of SIs. Semantic segmentation

[10] Adapted from "Formulation and Solution of the Problem of the Positions of a Mechanism with a Parallel–Series Structure Used in Surgery as an Alternative to the DA VINCI Robot", Veliev, and et al., 2019, Journal of Machinery Manufacture and Reliability, 48, 283-291, Copyright 2019 by "*Springer Nature*". Adapted with permission.

methods applied to robotic surgery images recognize occluded instruments as a single object when the SI locations are close or overlapping [101, 102]. Maintenance of the identity of each SI is critical for accurate determination of SI trajectory [103, 104]. The identity of SI is easily changed mainly when the SI goes out of the screen or is close to another SI.

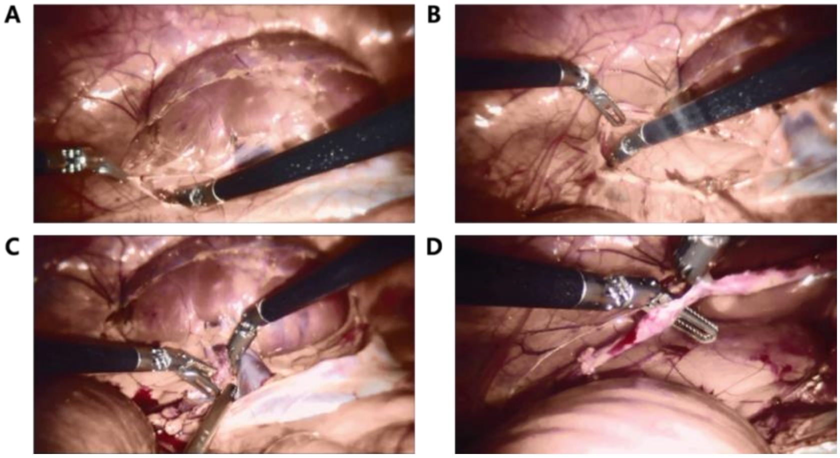


Figure 3.4 Examples of complex robotic surgery view. (A) Smoke (B) Variant illumination (C) Occlusion between surgical instruments (D) Occlusion between surgical instrument and tissue

The present study proposes a system that automatically and quantitatively assesses the surgical skill of a surgeon during robotic surgery by visual tracking of SIs using a deep learning method. The algorithm consists of two frameworks: instance segmentation for occlusion and tracking for maintaining types of SIs. This method was able to stably track the tip positions of SIs in patients with thyroid cancer undergoing robotic thyroid surgery with a bilateral axillo-breast approach (BABA) and in a BABA training model [105, 106]. The trajectory of the instruments enabled calculation of defined motion metrics [107], which were used to develop a system for quantitative assessment of surgical skill.

3.2 Methods

3.2.1 Study Design

We developed a deep learning-based tracking algorithm of multiple SIs to assess surgical skill in robotic surgery. Figure 3.5 shows an overview of the surgical skill assessment system used in robotic surgery. The system consists of two processes, the SI tracking algorithm and surgical skill assessment.

The SI tracking algorithm is a pipeline of deep learning-based methods involving an instance segmentation framework and a tracking framework, along with image processing methods to detect the tips of SIs and to recognize indicators (Figure 3.5A). The outputs of the instance segmentation framework were a bounding box and a mask of instruments on a surgical view. The results of the bounding box were input into the tracking framework, involving each SI frame by frame to maintain the type of instruments over time. The mask results were used to detect the positions of SI tips. To accurately determine the trajectory of each SI, it was necessary to detect the position of its tip, not its center [108]. An indicator recognition algorithm was applied to determine the moment of a laparoscopy usage and the status of an identified SI during robotic surgery. This prevented changes in laparoscopic views and errors due to immobile but present SIs in these views from being included in the trajectory.

Throughout the process of SI tracking, surgical skill was evaluated based on the acquired trajectory. Motion metrics [107] are quantitative indices, mainly related to the movement of SIs [25], in robotic surgical environments. Nine types of motion metrics were defined, with the metrics calculated based on SI trajectories (Figure 3.5B). In addition, surgical skill scores were determined by surgeons based on selected items related to SI motions in Object Structured Assessment of Technical Skills (OSATS) [37] and Global Evaluative Assessment of Robotic Surgery (GEARS) [38]. Finally, calculated motion metrics were used to develop a model predicting the surgical skill of novice, skilled, and expert robotic surgeons. This retrospective study was approved by the Institutional Review Boards of Seoul National University Hospital (IRB No. H-1912-081-1088).

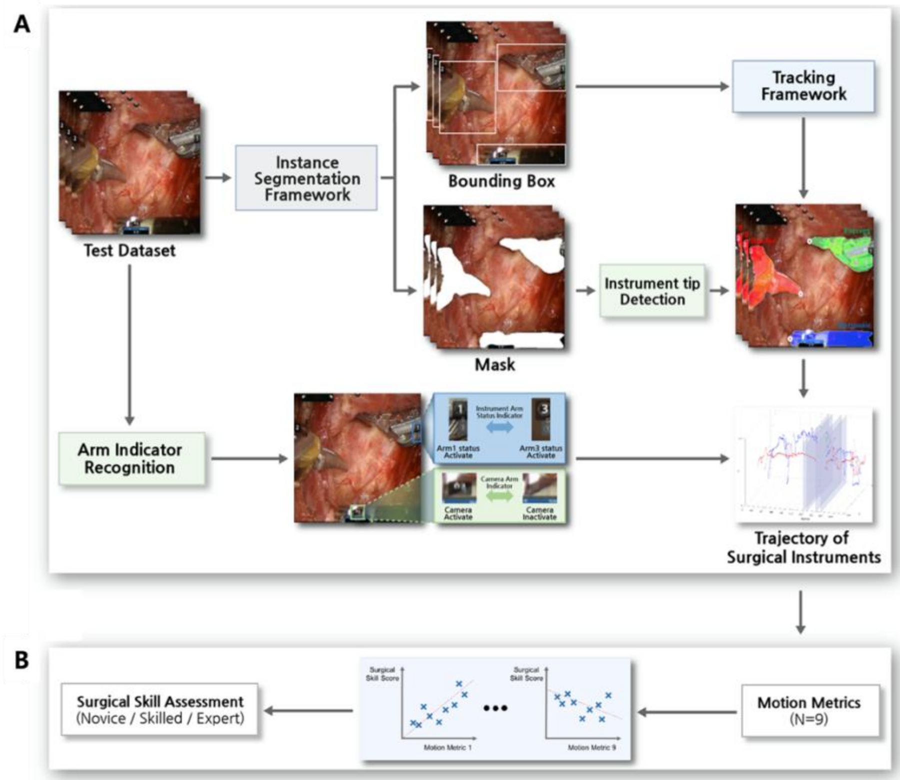


Figure 3.5 Overview of the surgical skill assessment system in robotic surgery. (A) Surgical instrument tracking algorithm. The pipeline consists of a deep learning-based instance segmentation framework and a tracking framework. Accurate trajectory of the surgical instruments was determined by surgical instrument tip detection and arm-indicator recognition. (B) Assessment of surgical skill. Motion metrics (e.g., instruments out of view) were calculated based on the acquired trajectory of surgical instruments and used to develop a surgical skill assessment system.

3.2.2 Dataset

The BABA to robotic thyroid surgery is a minimally invasive method used worldwide [106, 109, 110]. First, small incisions about 1 cm in size were placed on both sides of the axillae and the breast areolae, and the robot was docked to remove the thyroid gland. A view similar to that of traditional open thyroidectomy and the sophisticated arm movements of the robot provide surgical stability. A BABA training model enabling surgeons to practice has been developed [105]. The video datasets used are segments from the beginning to the locating of the recurrent laryngeal nerve (RLN) during thyroid surgery. Because injury to the RLN is a major complication of thyroid surgery, it is important to preserve RLN function during thyroid surgery [111].

Several types of daVinci surgical robots were used (S, Si, and Xi), along with four types of SIs: bipolar, forceps, harmonic, and cautery hook. The developed algorithm was applied to two types of surgical image. The first was a surgical image of a BABA training model developed for thyroid surgery training [105, 106]; subjects tested on this image included students, residents, and fellows. The second was a surgical image of a patient with thyroid cancer; subjects tested on this image included fellows and professors [112]. The data used to train spatial-temporal re-identification (ST-ReID) in the tracking framework consisted of 253 frames from patients (Table 3.1)

Test datasets consisted of 14 videos from the BABA training model and 40 videos from patients. Test video lengths ranged from 1121 to 40,621 frames, with a 23 fps. A detailed description of the test datasets is given in Table 3.2.

Table 3.1 Training datasets for the instance segmentation framework and spatial-temporal re-identification.

Training Dataset	No. of Videos	Total No. of Frames	Types of Surgical Instruments			
			Bipolar	Forceps	Harmonic	Cautery hook
BABA[11] training model (Instance Segmentation Framework [113])	10	84	158	82	-	-
Patients with thyroid cancer (Instance Segmentation Framework [113])	2	454	311	194	141	311
Public database [112] (Instance Segmentation Framework [113])	8	1,766	1.451	1,351	-	-
Patients with thyroid cancer (ST-ReID[12] [114])	3	253	99	77	81	58

[11] BABA, bilateral axillo-breast approach

[12] ST-ReID, spatial-temporal re-identification

Table 3.2 Description of test datasets. The test dataset consisted of 14 videos of the axillo-breast approach (BABA) to thyroid surgery and 40 videos of patients. The number of scenes is the number of segmented videos at the point of time when laparoscopy was used.

Test dataset	Video no.	No. of scenes	Total no. of frames	Video no.	No. of scenes	Total no. of frames
BABA training model	1	6	11,087	8	1	7,639
	2	7	6,978	9	6	10,687
	3	11	7,638	10	3	10,744
	4	4	10,086	11	1	10,763
	5	6	7,323	12	1	7,298
	6	2	6,140	13	2	11,174
	7	2	11,285	14	2	7,142
	Total	54	125,984	-	-	-
Patients with thyroid cancer	1	2	2,869	21	19	22,748
	2	6	16,489	22	5	3,570
	3	2	7,262	23	4	8,375
	4	8	32,846	24	16	2,832
	5	3	6,260	25	8	11,227
	6	8	15,401	26	5	8,483
	7	6	18,587	27	5	1,875
	8	7	6,909	28	11	9,212
	9	7	18,249	29	10	13,359
	10	5	11,967	30	21	40,621
	11	2	8,202	31	2	2,831
	12	3	4,710	32	4	5,587
	13	5	8,407	33	5	3,452
	14	1	2,357	34	3	3,511
	15	8	9,787	35	16	7,140
	16	7	11,046	36	22	20,494

17	6	6,009	37	9	9,209
18	2	1,121	38	13	3,584
19	5	11,236	39	4	3,307
20	5	4,613	40	1	2,140
Total	281	387,884	-	-	-

3.2.3 Instance Segmentation Framework

Unlike semantic segmentation methods applied to robotic surgery images [23, 115], the instance segmentation method can separate occluded instruments during the first stage, followed by semantic segmentation during the second stage.

The instance segmentation framework used is Mask R-CNN [113], which in order, consists of an RPN [116], region of interest (ROI) classifier with bounding box regressor, and semantic segmentation networks as shown in Figure 3.6. The backbone of the CNN used are ResNet101 [117] and feature pyramid network (FPN) [118]. Second, RPN scans over the backbone feature maps, called anchors which are 256 different sizes and aspect ratios, covering images as much as possible. Furthermore, non-maximum suppression (NMS) is applied, so the box with the highest confidence score was selected, and if the intersection over union (IoU) with the corresponding box was above the threshold of 0.9, it was finally selected. Next, region of interest (ROI) pooling layer resizes a feature map to a fixed size by bilinear interpolation. In addition, ROI classifier, softmax, classifies surgical instruments (foreground) and background, and bounding box regressor refines the coordinates of bounding boxes. Finally, the last CNN layers segment 28 x 28 soft mask to resized binary mask.

In the training phase, the loss function of Mask R-CNN, defined as a multi-task loss, can be expressed as equation (1):

$$L = L_{cls} + L_{box} + L_{mask} \tag{1}$$

The classification loss L_{cls} and bounding box loss L_{box} are identical to those defined for Fast R-CNN [116], whereas L_{mask} is an average binary cross-entropy applied to a per-pixel sigmoid regardless of the appearance or type of SIs. Adam optimizer with a learning rate of 0.001 was used [68]. In addition, augmentation techniques, such as rotation, flip, and brightness adjustment were applied.

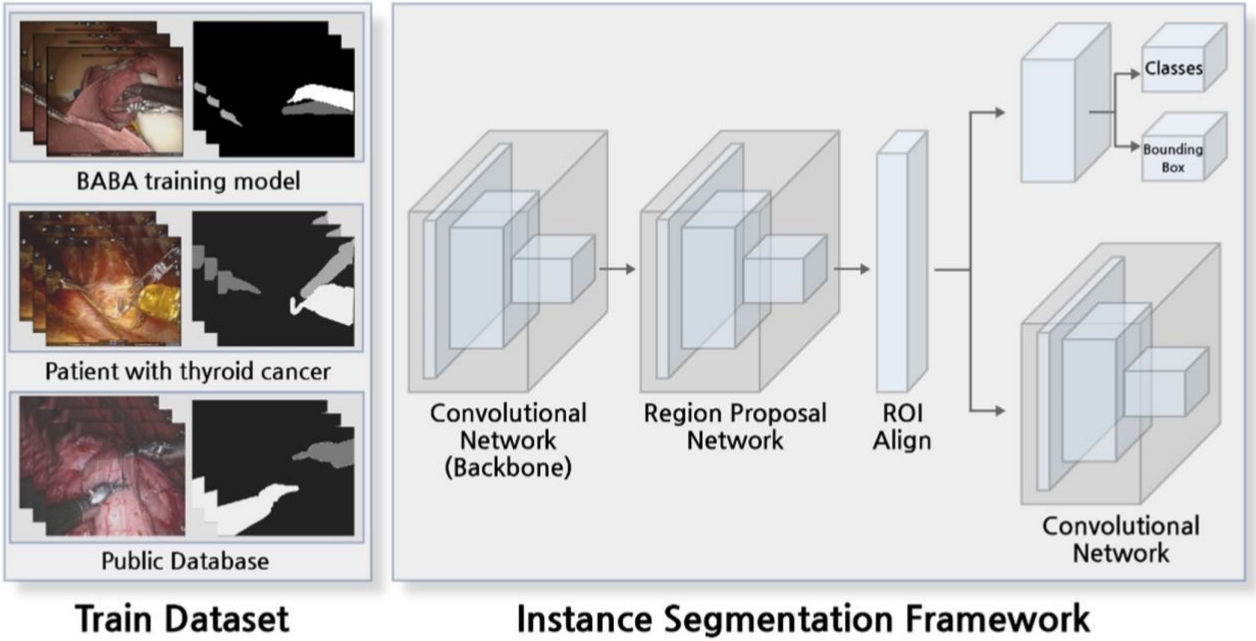


Figure 3.6 The overview of the instance segmentation framework. The framework was trained with three types of training datasets: the BABA training model, patients, and a public database.

3.2.4 Tracking Framework

The positions of the SIs determined by results of the instance segmentation framework, the bounding boxes, must be assigned to the next frame of the same SIs. The tracking framework was designed to associate the identity of an SI to the next identity of that SI and maintain these associations over frames as shown in Figure 3.7. The framework consists of a cascade structure, a tracker, and re-identification method as shown in Figure 3.8.

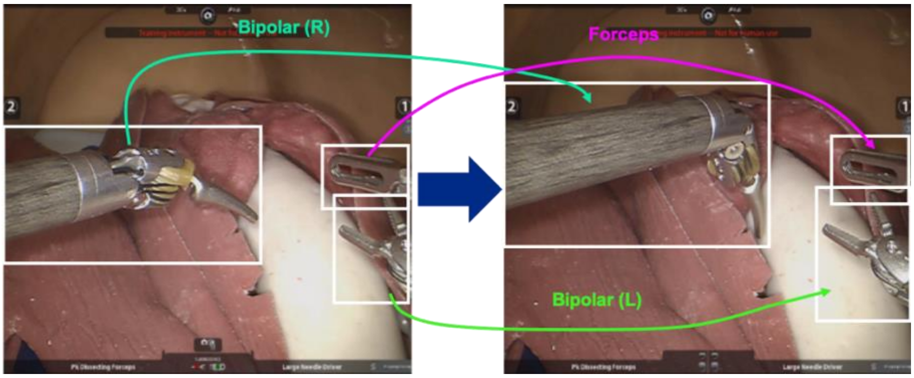


Figure 3.7 Concept of tracking: Association of objects over frames.

3.2.4.1 Tracker

The tracker used in this study was a deep simple online and realtime tracker (deep SORT) [104] which associated target SIs in consecutive video frames using spatial and temporal information (Figure 3.9). The algorithm operated in the following order. The final bounding box was selected from among the bounding box candidates through a non-maximum suppression method [104] as a result of the instance segmentation framework. Next, the Kalman filter using time information [119] and the intersection-over-union (IOU) using spatial information were applied to associate the identity of SIs that move over time. A Hungarian algorithm was used for optimization of the final selection in association with SIs [120].

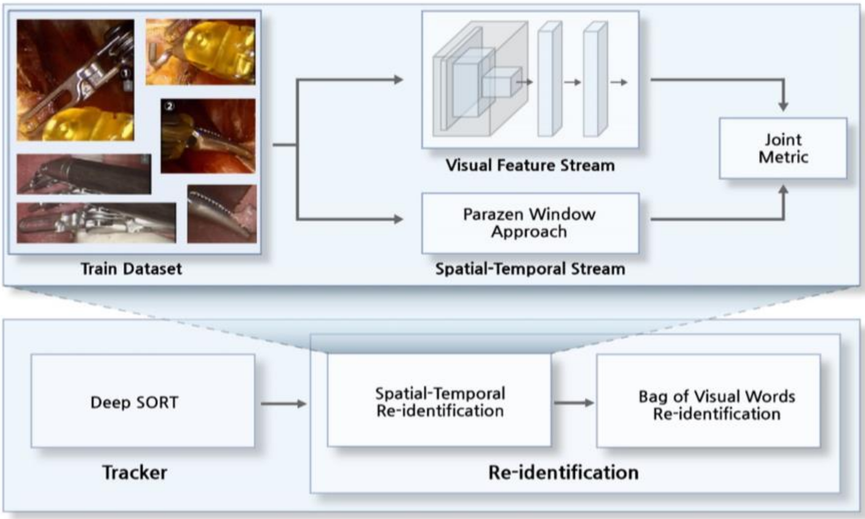


Figure 3.8 The overview of the tracking framework. The tracking framework consists of a tracker and a sequence of re-identification algorithms. The spatial-temporal re-identification algorithm was trained with bounding boxes of all types of surgical instrument.

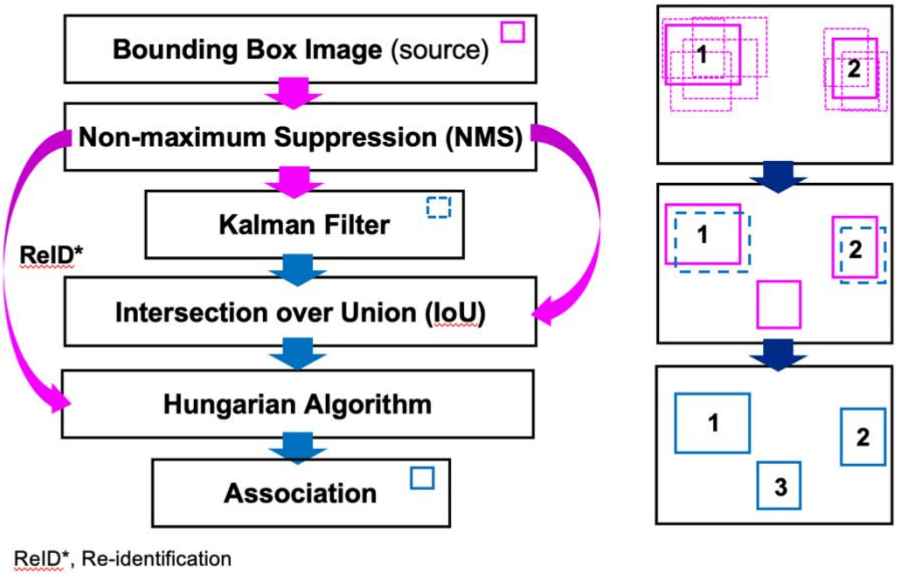

Bounding Box Image (source)
Non-maximum Suppression (NMS)
Kalman Filter
ReID*
Intersection over Union (IoU)
Hungarian Algorithm
Association
ReID*, Re-identification
1
2
1
2
1
2
3

Figure 3.9 Block diagram of Deep SORT

3.2.4.2 Re-Identification

Re-identification (ReID) was applied to the result of Deep SORT because the existing identity of an SI can change when an SI moves out of view or when SIs cross in close proximity. In addition, the maximum number of SIs that appear on the robot surgery view was set at three, thus limiting the number of SIs.

In the proposed ReID method, offline and online learning methods were applied sequentially. Spatial temporal re-identification (ST-ReID) [114] is an offline learning method that trains all types of SIs in advance using spatial and temporal information. This method consists of three sub-modules, a visual feature stream, a spatial-temporal stream, and a joint similarity metric, the latter of which integrates two streams of information and a fast Histogram-Parazen (HP) to approximate a complex spatial-temporal probability distribution. Histograms were smoothed with the HP method using equation (2):

$$p(y = 1|k, c_i c_j) = \frac{1}{Z}\sum_l \hat{p}(y = 1|l, c_i c_j)K(l - k) \qquad (2)$$

where k indicates the kth bin of a histogram, c denotes the camera IDs, K is a gaussian function, and $Z = \sum_k \hat{p}(y = 1|l, c_i c_j)$ is a normalization factor.

Following spatial-temporal re-identification (ST-ReID), the bag of visual words (BOVW) which is online learning method [121] was applied. The moment the prior method predicted changes in identity, the visual features of SIs during certain previous frames were trained and reflected in these changes. At the moment the ST-ReID predicted changes in identity, the visual features of an SI extracted by ORB descriptor [122] for fewer than 10 previous frames were trained, and the identity of the surgical instrument (SI) was predicted using an support vector machine.

3.2.5 Surgical Instrument Tip Detection

The SI tip detection algorithm was applied because the tip position more accurately reflects the movement of the SI than a detection algorithm which yields the center of SI bounding box [108]. SI tips were detected from the binary SI mask, which resulted from the instance segmentation framework. The starting point was determined by considering the number of SI masks contacted among eight defined sections a certain distance from the edge of the view. A skeletonization algorithm was applied to the SI mask [123], and the position of the SI tip in the skeletonized SI was determined by calculating the longest accumulated distance from the starting point. Finally, the kalman filter was applied to minimize outliers [100, 119]. Figure 3.10 describes the detail of the procedure of the tip of surgical instruments detection.

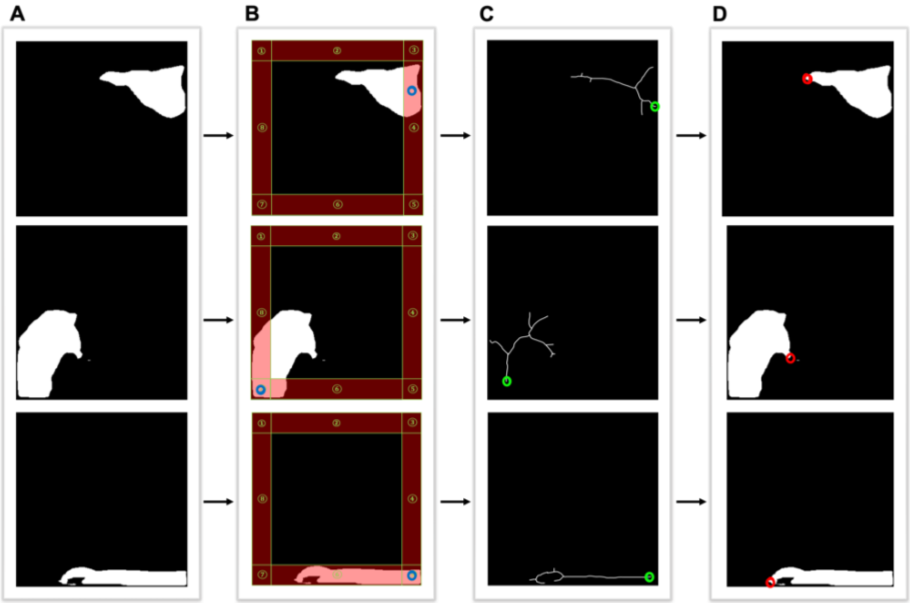


Figure 3.10 Procedure of the tip of surgical instruments detection. (A) Surgical instrument mask from the instance segmentation framework. (B) Starting point detection in the mask from edges of the view. The area located at a certain distance from the edge, and the area at which the mask overlaps was determined. The starting point (blue) was based on the number of contacted sections at a certain distance from the edge of the view. (C) Application of the skeletonization algorithm to the mask to determine the main skeleton. After calculating the skeleton, update the position of the starting point (green) to the

nearest position in the skeleton. (D) Determination of the tool tip position (red) within the skeleton by calculating the longest accumulated distance from the starting point, and kalman filter was applied.

3.2.6 Arm-Indicator Recognition

The arm indicators that could have affected the trajectory consisted of instrument arm status and camera arm indicators. The instrument arm status indicator on both sides of the screen indicated the SI currently in use. Therefore, these indicators reflected the movement of two or fewer SIs actually being used rather than the movement of the SI that appeared in the robotic surgery view. Recognition of the camera arm indicator confirmed the movement of the laparoscope during the operation. The appearance of the camera arm indicator on the robotic surgery view indicated movement of the laparoscope; however, movement of the screen may have incorrectly indicated movement of the SI. Although varying according to the type of surgical robot, the positions of both indicators were fixed on the view and appeared when an event occurred. To recognize the arm-indicator, template matching [124] was applied to the robotic surgery view. Because the shape and position of the indicators were fixed, the template of each arm-indicator was stored in advance.

3.2.7 Surgical Skill Prediction Model

Two surgeons reviewed recorded videos of surgeons being trained using the BABA training model and of surgeons performing thyroid surgery on patients with thyroid cancer [125, 126]. Parts of items and related motion metrics in OSATS and GEARS were scored [105, 127]. The defined items included time and motion, instrument handling, and flow of operation and forward planning in OSATS (Table 3.3). The item of respect for tissue is exclude, because the proposed method is recognition algorithm for surgical instruments (SIs) and not the algorithm for surrounding objects such as tissue, it cannot be addressed in this study. However, since the video segment used in this experiment is a process until the search for the main structure, recurrent laryngeal nerve (RLN), it is appropriate to exclude the item because it is an experiment in which a continuous contact with tissue is inevitable. In addition, the items of knowledge of instruments, use of assistants, and knowledge of the specific procedure were excluded, for they are not suitable for video analysis [127].

Next, the items of bimanual dexterity, efficiency, and robotic control are selected in GEARS (Table 3.4) for this study. Since the proposed SI tracking algorithm was applied to the 2-dimensional image, the item of depth perception was excluded, and the force sensitivity was excluded because it did not recognize objects other than SIs for the same reason as OSATS. Also, the items of autonomy and use of third arm were excluded, for they were not suitable for video analysis [128]. Each item was scored from 1 to 5 with a total of 15 grades.

Table 3.3 Description of object structured assessment of technical skills (OSATS) with relevance to motion metrics.

No.	OSATS[13] [37]	1	2	3	4	5	Relevance to motion metrics
1	Respect for Tissue	Frequently used unnecessary force on tissue or caused damage by inappropriate use of instruments		Careful handling of tissue but occasionally caused inadvertent damage		Consistently handled tissues appropriately with minimal damage	X
2	Time and Motion	Many unnecessary moves		Competent use of instruments although occasionally appeared stiff or awkward		Economy of movement and maximum efficiency	O
3	Instrument Handling	Repeatedly makes tentative or awkward moves with instruments		Competent use of instruments although occasionally appeared stiff or awkward		Fluid moves with instruments and no awkwardness	O

[13] OSATS, Object Structured Assessment of Technical Skills

4	Knowledge of Instruments	Frequently asked for the wrong instrument or used an inappropriate instrument	Knew the names of most instruments and used appropriate instrument for the task	Obviously familiar with the instruments required and their names	X
5	Use of Assistants	Consistently placed assistants or failed to use assistants	Good use of assistants most of the time	Strategically used assistant to the best advantage at all times	X
6	Flow of Operation and Forward Planning	Frequently stopped operating or needed to discuss next move	Demonstrated ability for forward planning with steady progression of operative process	Obviously planned course of operation with effortless flow from one move to the next	O
7	Knowledge of Specific Procedure	Deficient knowledge. Needed specific instruction oat most operative steps	Knew all important aspects of the operation	Demonstrated familiarity with all aspects of the operation	X

Table 3.4 Description of global evaluative assessment of robotic surgery (GEARS) with relevance to motion metrics.

No.	GEARS[14] [38]	1	2	3	4	5	Relevance to motion metrics
1	Depth Perception	Constantly overshoots target, wide swings, slow to correct		Some overshooting or missing of target, but quick to correct		Accurately directs instruments in the correct plane to target	X
2	Bimanual Dexterity	Use only one hand, ignores non-dominant hand, poor coordination		Uses both hands, but does not optimize interactions between hands		Expertly uses both hands in a complementary way to provide best exposure	O
3	Efficiency	Inefficient efforts; many uncertain movements; constantly changing focus or persisting without progress		Slow, but planned movements are reasonably organized		Confident, efficient and safe conduct, maintains focus on task, fluid progression	O
4	Force Sensitivity	Rough moves, tears tissue, injures nearby structures, poor control, frequent suture breakage		Handles tissues reasonably well, minor trauma to adjacent tissue, rare suture breakage		Applies appropriate tension, negligible injury to adjacent structures, no suture breakage	X

[14] GEARS, Global Evaluative Assessment of Robotic Surgery

5	Autonomy	Unable to complete entire task, even with verbal guidance	Able to complete task safely with moderate guidance	Able to complete task independently without prompting	X
6	Robotic Control	Consistently does not optimize view, hand position, or repeated collisions even with guidance	View is sometimes not optimal. Occasionally needs to relocate arms. Occasional collisions and obstruction of assistant.	Controls camera and hand position optimally and independently. Minimal collisions or obstruction of assistant	O
7	Use of Third Arm	Consistently does not use it, or does not use it well when required, even with verbal guidance	Mostly uses 3rd arm in a safe and efficient manner with moderate guidance	Consistently uses 3rd arm in a safe and efficient manner without prompting	X

Based on the acquired trajectories, motion metrics, mainly related movements of SIs, were used to develop a surgical skill prediction model. Seven metrics associated with motion were included [94, 107, 129]: time to completion of surgery, instruments out of view, instrument collision, economy of motion, average speed, number of movements, and economic factors. Two additional metrics related to the robotic surgery environment, surgical instrument changes and laparoscopy usage, were newly defined for this study (Table 3.5).

The surgical skill prediction models were developed using these nine calculated motion metrics as well as ground truth from OSATS and GEARS scores. The total number of tested videos was 54, with these datasets divided into 40 training and 12 test sets. Surgical skill prediction models were developed using machine learning methods, a linear classifier, a support vector machine (SVM), and random forest, with the model predicting three groups consisting of novice, skilled, and expert surgeons.

In the training process, five-fold cross validation was applied, and the class imbalance issue was solved by applying the synthetic minority over-sampling technique (SMOTE) [130]. SVM used the Gaussian kernel, with the external hyperparameter optimized through training being a regularization parameter. Additionally, the random forest was trained based on the Gini impurity, with the external hyperparameters optimized through training being the number of trees and the maximum depth of the tree. The selected hyperparameters were trained and fine-tuned during 500 epochs.

3.3 Experiments and Results

3.3.1 Performance of Instance Segmentation Framework

Figure 3.11 shows the qualitative results of the instance segmentation framework in the BABA training model and in a patient with thyroid cancer. This result shows that even when occlusion occurs between surgical instruments, it can be recognized by each surgical instrument. The process of train and validation loss are shown in Figure 3.12 A-B.

Table 3.5 Description of motion metrics. Motion metrics were defined in reference to the robotic surgical environment, and consist primarily of movements of surgical instruments and numbers of laparoscopes.

No.	Motion metric	Description
1	Time to complete (s) [94]	Total time from the beginning to end of all surgical procedures
2	Instruments out of views (s) [94]	Total distance traveled by all instruments when not in view
3	Instrument collision (n) [94]	Number of times one instrument collided with another instrument
4	Economy of motion (mm) [94]	Total distance traveled by instruments
5	Average speed (mm/s) [107]	Rate of change of the instrument's position in the image
6	Number of movement (n) [129]	Number of times beyond an acceleration of tolerance threshold
7	Economy of Area (-) [107]	Relationship between the maximum image area occupied by the instrument and the total distance
8	Surgical instrument change (n)	Number of surgical instrument type changes
9	Laparoscopy usage (n)	Number of times of laparoscopy usage

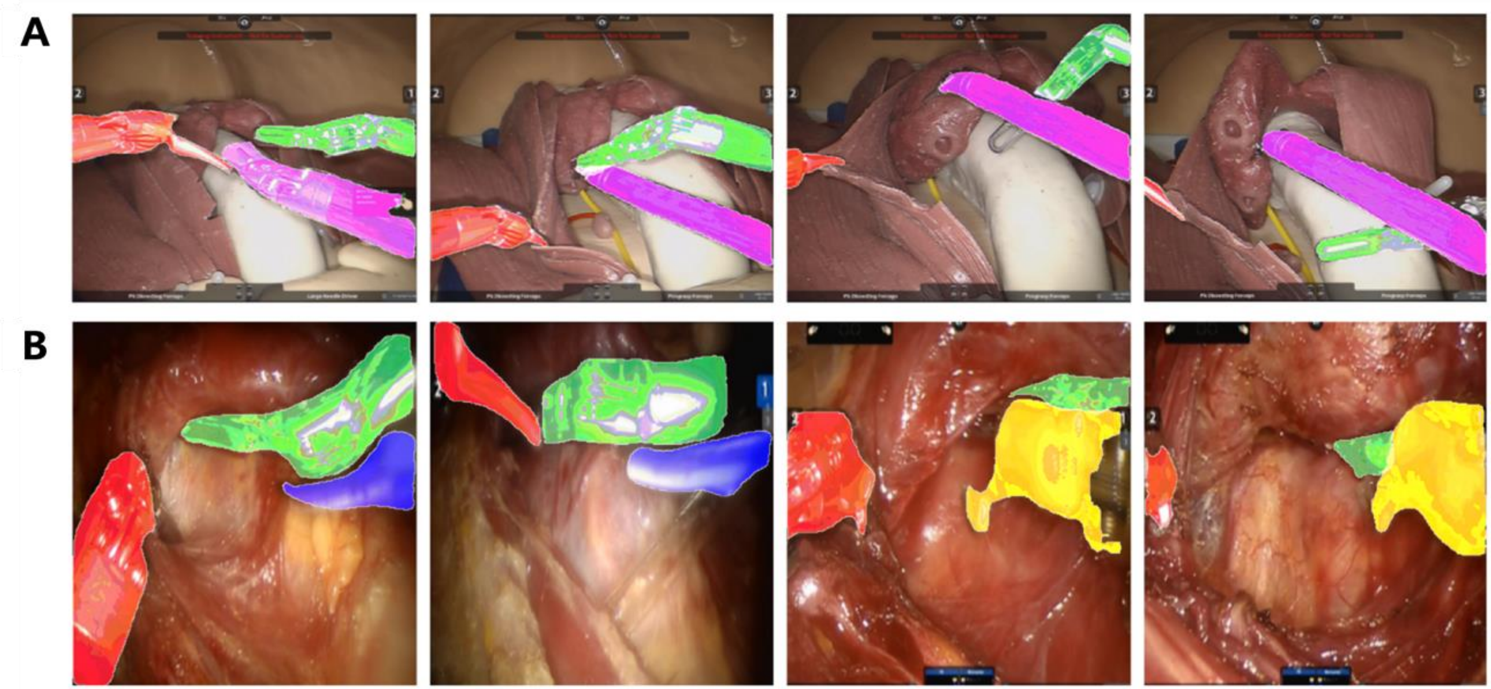


Figure 3.11 Qualitative results of the instance segmentation framework. Recognition of occlusion between surgical instruments located close together or overlapping (red: bipolar (left); pink: bipolar (right); green: forceps; blue: harmonic; yellow: cautery hook) (A) Sample results applied to the BABA training model. (B) Sample results applied to patients.

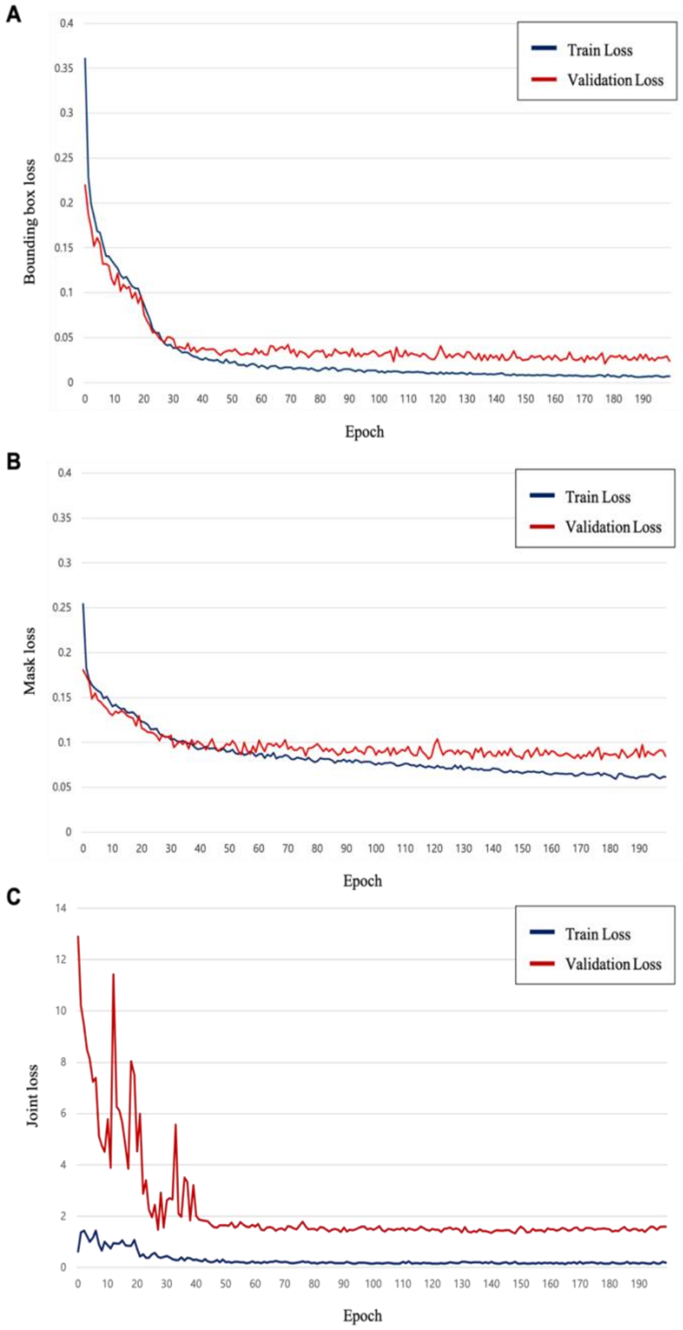


Figure 3.12 Plots of model loss on the training and validation datasets. (A) Loss of bounding box in the instance segmentation framework. (B) Loss of mask in the instance segmentation framework. (C) Loss of spatial-temporal re-identification.

3.3.2 Performance of Tracking Framework

Cumulative Matching Characteristics (CMC) [131], shown in equation (3), were used to evaluate the proposed tracking method at the moment the identity of SI predicted by the previous deep SORT algorithm was not maintained. Table 3.6 shows the comparative performance of ReID methods. Before applying the ReID methods, when only Deep SORT was applied, the accuracy of applying the ReID method was measured by setting the ratio of the identity of SIs to 0% as a reference point. The evaluation metric ranked at most three types of SI samples according to their distances to the query. The combination of ST-ReID with BOVW-ReID showed accuracy 93.3% with the BABA training model and 88.1% in patients with thyroid cancer. The process of train and validation loss are shown in Figure 3.12 C.

$$Accuracy_1 = \begin{cases} 1 & if\ top1\ ranked\ SI\ samples\ contatin\ the\ query\ identity \\ 0 & otherwise \end{cases} \tag{3}$$

Table 3.6 Comparative performance of re-identification methods.

ReID[15] Method	BABA[16] Training Model (Rank-1)	Patients with Thyroid Cancer (Rank-1)
BOVW[17] [121]	68.3%	57.9%
ST-ReID[18] [114]	91.7%	85.2%
BOVW [121] + ST-ReID [114]	93.3%	88.1%

3.3.3 Evaluation of Surgical Instruments Trajectory

Figure 3.13 shows the trajectory of multi-SIs tip, as determined by the proposed tracking algorithm. The differences between the algorithm-based determination of the tip position and the ground truth, labeled at 2 frames per second (23 frames), were determined. The root mean square error (RMSE) averaged 2.83mm for the BABA training model, and 3.75mm in patients with thyroid cancer. The unit of distance that each SI moved was converted from pixels to millimeters because the width and the height of each image were dependent on the type of da Vinci robot used. Thus, depending on the degree of magnification of the laparoscope, errors may have occurred when calculating the movement of the actual SIs. For unit conversion, the thickness of the surgical instrument was measured in advance (8 mm), and the thickness shown

[15] ReID, re-identification

[16] BABA, bilateral axillo-breast approach

[17] BOVW, bag of visual words

[18] ST-ReID, spatial-temporal re-identification

on the first screen was measured in pixels units. Therefore, through proportional relationships, the motion of each SI in pixels was converted to millimeters in all surgical images [132].

This system also measured whether the end position of the SI predicted by the algorithm was within 1, 2, and 5 mm of the end position of the SI on the screen [24]. True positive and false positive results were obtained using a confusion matrix (Figure 3.14). Therefore, area under the curve (AUC) could be calculated by plotting a receiver operating characteristic (ROC) curve using true positive and false positive rates.

The mean AUC for errors within 1, 2, and 5 mm were 0.73, 0.83, and 0.92, respectively, in the BABA training model and 0.69, 0.76, and 0.84, respectively, in patients with thyroid cancer. Finally, Pearson's correlation analysis, performed to assess the similarity between predicted trajectories and ground truth, showed that these trajectories were 0.93 (*x*-axis) and 0.91 (*y*-axis) in the BABA training model and 0.89 (*x*-axis) and 0.86 (*y*-axis) in thyroid cancer patients (Table 3.7).

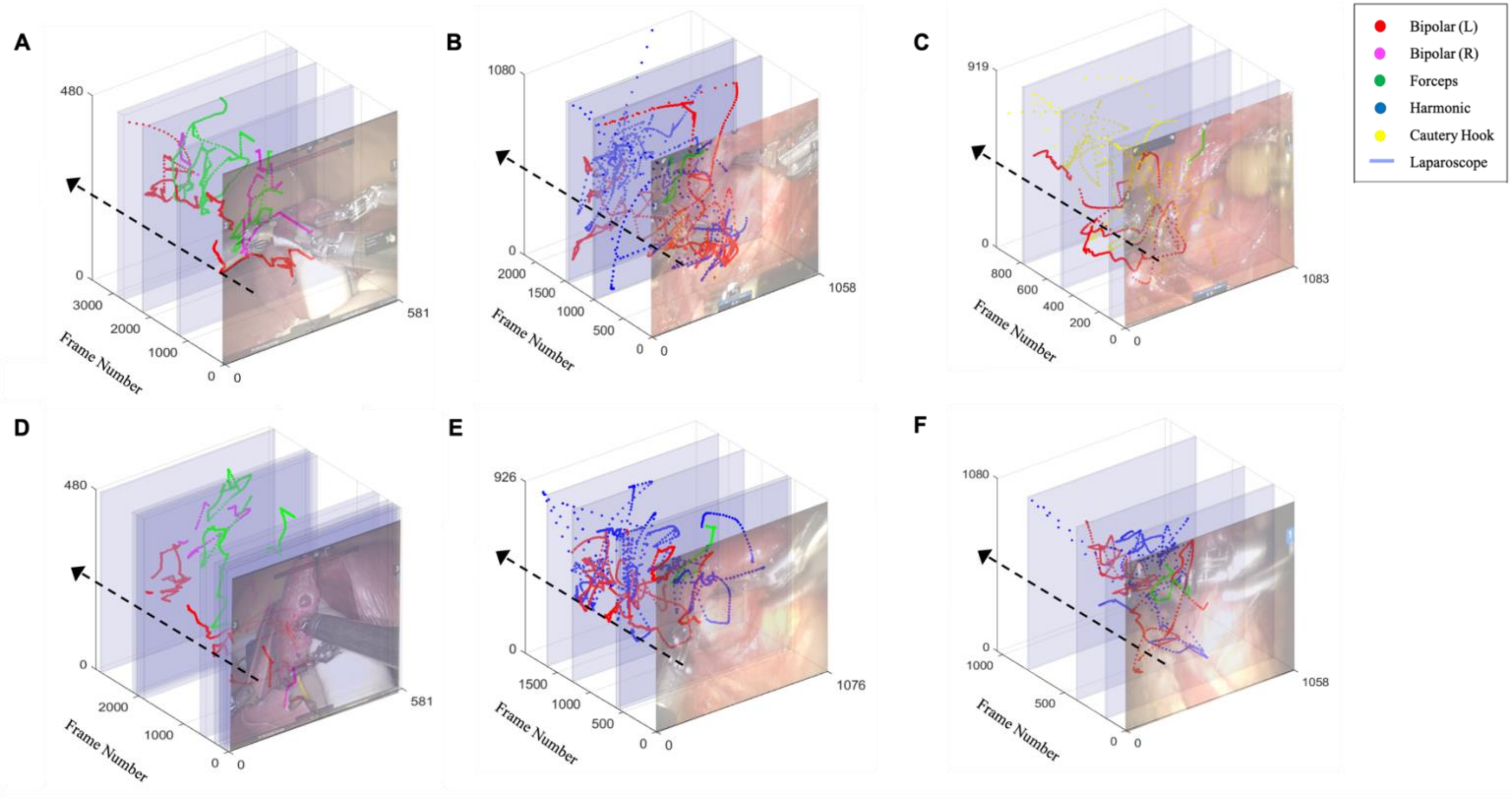


Figure 3.13 Trajectory of multi-surgical instruments tip. Each color represents a type of surgical instrument, and the blue area represents the duration of laparoscopy. (A,D) Trajectory of novice surgeons (B,E) Trajectory of skilled surgeons (C,F) Trajectory of expert surgeons.

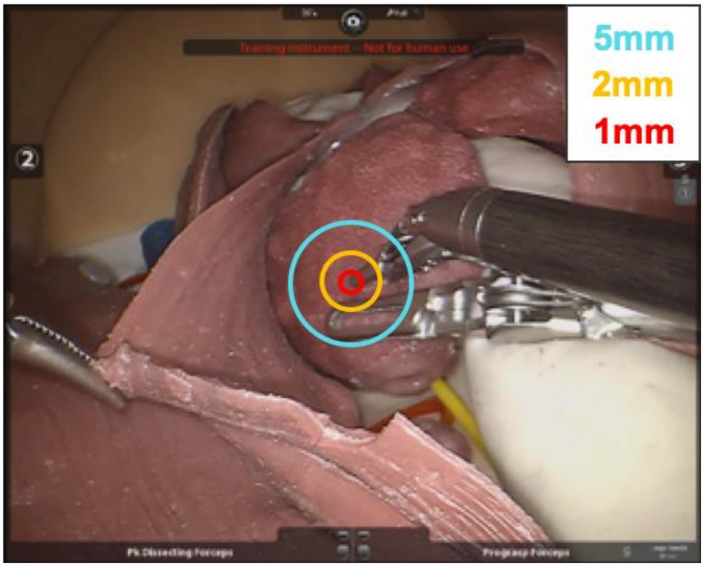


Figure 3.14 Performance evaluation method between algorithm and ground truth. Red, yellow and cyan circles are 1, 2, and 5mm respectively. In our experiments, if the location of a prediction is within the circles, we consider it as true positive, and otherwise, we consider it as false positive.

3.3.4 Evaluation of Surgical Skill Prediction Model

The OSATS and GEARS scores of the two surgeons showed an intra-class correlation coefficient (ICC) of 0.711 with OSATS and 0.74 with GEARS. Each motion metric item was normalized to the operation time and then to the metrics.

The performance of linear, SVM, and random forest surgical skill prediction models were compared. The models were optimized by hyper parameter tuning, with the random forest showing the highest accuracy. The random forest model had the highest performance and accuracy of 83% with OSATS and 83% with GEARS. Figure 3.15 shows a comparison of the performance of these surgical skill prediction models. In addition, the relative importance of motion metrics was analyzed in OSATS and GEARS. As shown in Figure 3.16, the most relative important metrics in both OSATS and GEARS was economy of motion, followed by the instrument out of view.

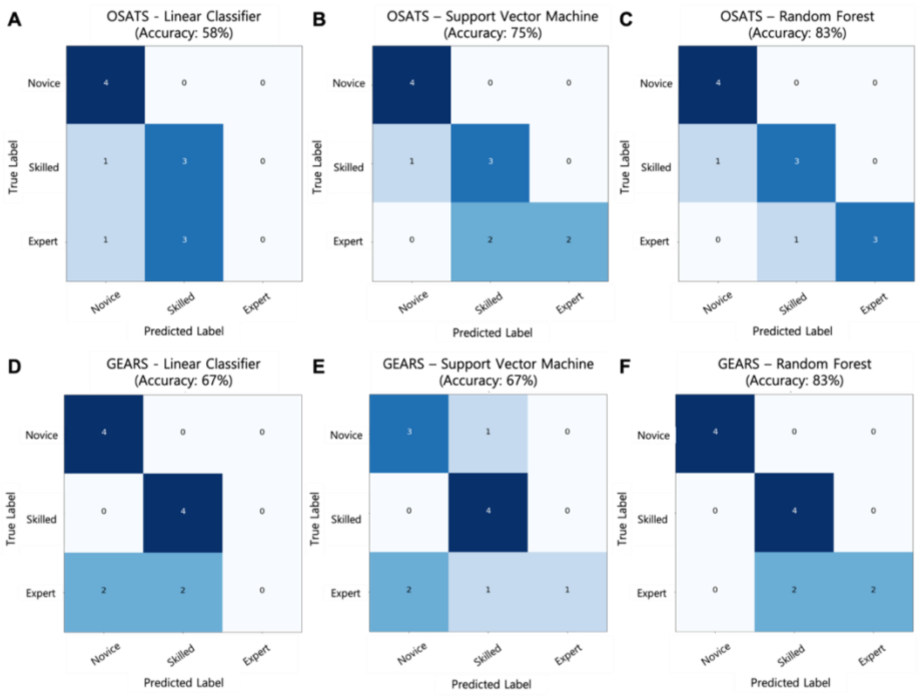


Figure 3.15 Comparison of the performance of surgical skill prediction models and parts of items in Object Structured Assessment of Technical Skills (OSATS) and Global Evaluative Assessment of Robotic Surgery (GEARS) with a confusion matrix. The test dataset consisted of four novice, four skilled, and four expert surgeons. (A-C) Confusion matrix results of models using the OSATS. (A) Linear classifier; (B) support vector machine; and (C) random forest. (D–F) Confusion matrix results of models using the GEARS. (D) Linear classifier; (E) support vector machine; and (F) random forest.

Table 3.7 Comparative performance of the surgical instruments tip detection. Evaluation methods included determinations of average are root mean square error (RMSE; mm), average area under the curve (AUC; 1, 2, and 5mm), and average Pearson correlation coefficient (x-axis and y-axis) between tip positions determined by the algorithm and ground truth.

Test Dataset (No. of Videos)	No. of Frames	RMSE[19] (mm)	AUC[20] (1mm)	AUC (2mm)	AUC (5mm)	Pearson-r[21] (x-axis)	Pearson-r (y-axis)
BABA[22] Training Model (n = 14)	125,984	2.83 ± 1.34	0.73 ± 0.05	0.83 ± 0.02	0.92 ± 0.02	0.93 ± 0.02	0.91 ± 0.04
Patients with Thyroid Cancer (n = 40)	387,884	3.7 ± 2.29	0.69 ± 0.04	0.76 ± 0.06	0.84 ± 0.03	0.89 ± 0.03	0.86 ± 0.03
Average (n = 54)	513,868	3.52 ± 2.12	0.7 ± 0.05	0.78 ± 0.06	0.86 ± 0.05	0.9 ± 0.03	0.87 ± 0.04

[19] RMSE, root mean square error

[20] AUC, area under the curve

[21] Pearson-r, pearson correlation coefficient

[22] BABA, bilateral axillo-breast approach

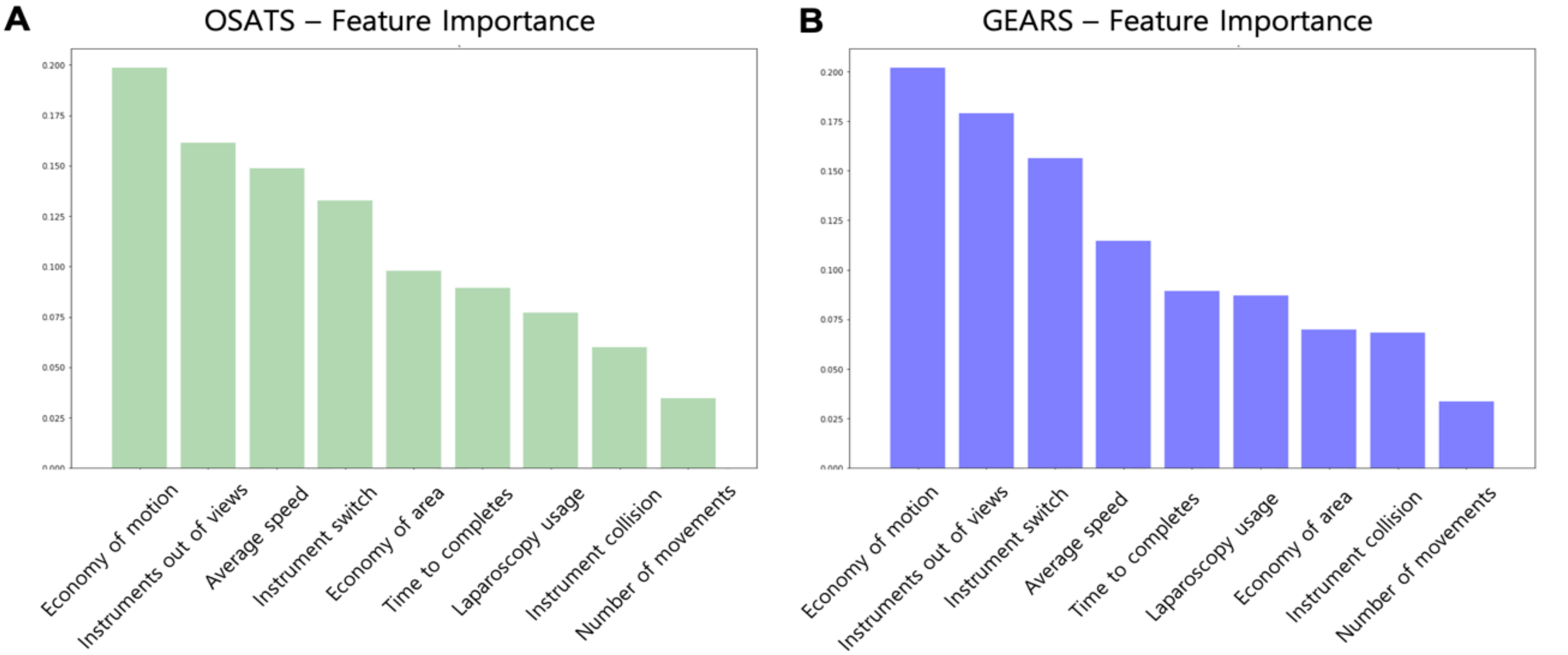


Figure 3.16 Relative importance of motion metrics in surgical skill prediction model. (A) Importance of motion metrics in Object Structured Assessment of Technical Skills (OSATS). (B) Importance of motion metrics in Global Evaluative Assessment of Robotic Surgery (GEARS).

3.4 Discussion

3.4.1 Research Significance

To the best of our knowledge, this is the first deep learning-based visual tracking algorithm developed for a quantitative surgical skill assessment system. Conventional methods of evaluating surgical skill such as OSATS [37] and GEARS [38], were based on assessments of recorded videos during robotic surgery from 0 to 30 grades. Because SI movements are associated with surgical skill, the newly proposed quantitative assessment method used a tracking algorithm to determine the trajectories of multiple SIs, showing an accuracy of 83% when compared with conventional methods.

Previously described SI tracking algorithms are limited by occlusion among different and multiple SIs being recognized as a single SI [101, 102]. SI identity cannot be maintained over frames because SIs have similar appearances, especially when only parts are visible [133, 134]. The proposed method overcomes occlusion using an instance segmentation framework and overcomes identity maintenance using a tracking framework. Accurate determination of SI trajectories enables the calculation of motion metrics and the quantitative evaluation of surgical skill.

The SI tracking algorithm was developed based on robotic surgical environments. In this study, four types of SIs were used, but if only the shaft of the SI appeared on the surgical screen, it could not be discerned, thus we approached the binary classification problem that distinguishes the SI (foreground) from the background. In addition, SI may be difficult to discern when it is covered by tissues or when only a part is visible during surgery [32]. Therefore, to minimize errors resulting from segmentation, the tracking algorithm used temporal information to determine the type of SI. The described tracking algorithm is typically used in a tracking framework to track pedestrians [104, 114]. The maximum number of SIs viewed during robotic surgery is three, limiting the number of objects recognized by the proposed algorithms. An arm-indicator recognition algorithm was applied to reflect a robotic surgery environment in which an SI appears but does not actually move. Specifically, the instrument arm status indicator provides information about the two activated SIs in use, with the camera arm indicator determining the moment the laparoscope was moved, preventing errors resulting from the trajectory of the immobile SI.

Our findings also confirmed that the four most important metrics in OSATS and GEARS were the same: economy of motion, instruments out of view, average speed, and instrument switch. The video datasets used in this study were video segments from the beginning of surgery to the locating of the RLN during thyroid surgery. Therefore, the relative importance of the motion metrics may differ depending on surgical sites and tasks.

3.4.2 Limitations

This study had several limitations. First, the proposed system was applied to video sets of training model and patients with thyroid cancer who underwent BABA surgery. It is necessary to verify the effectiveness of the proposed system using various surgical methods and surgical areas.

Second, we could not directly compare the performances of the kinematics and proposed image-based methods because access to the da Vinci Research Interface is limited, allowing most researchers only to obtain kinematic raw data [27]. However, previous studies have reported that the kinematics method using da Vinci robot had an error of at least 4 mm [135]. Direct comparison of performance is difficult because the surgical images used in the previous study and in this study differed. However, the average RMSE of the proposed image-based tracking algorithm was 3.52 mm, indicating that this method is more accurate than the kinematics method and that the latter cannot be described as superior.

The performance of the current method with the previous visual method could not be directly compared because no similar study detected and tracked the tip coordinates of the SIs. However, studies have used deep learning-based detection methods to determine the bounding boxes of the SIs and to display the trajectory of the center points of these boxes [22, 32, 108]. Nevertheless, because this approach could not determine the specific locations of the SIs, it cannot be considered an accurate tracking method intuitively. Comparison of the quantitative performance of the proposed method and other approaches is important, making it necessary to compare different SI tracking methods.

Third, because SIs are detected on two-dimensional views, errors may occur due to the absence of depth information. In the future, methods are needed to utilize three-dimensional information based on stereoscopic matching of left and right images during robotic surgery [136, 137].

Fourth, because the proposed method is a combination of several algorithms, longer videos can result in the accumulation of additional errors, degrading the performance of the system. Thus, in particular, it is necessary to train additional negative examples with the instance segmentation framework, which is the beginning of the pipeline. For example, gauze or tubes on the robotic surgery view can be recognized as SIs (Figure 3.17).

Finally, because errors from re-identification in the tracking framework could critically affect the ability to determine correct trajectories, accurate assessment of surgical skill requires manual correction of errors (Figure 3.18).

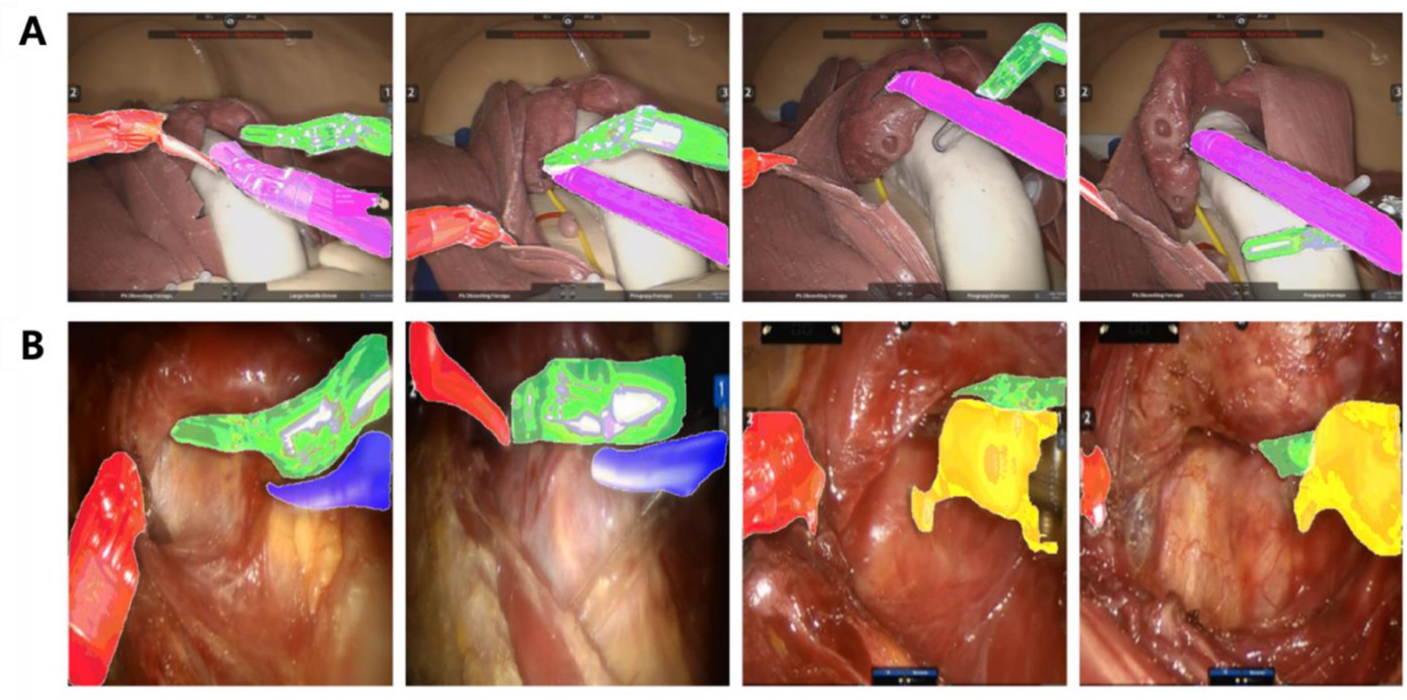


Figure 3.17 Errors in the instance segmentation framework. (Top: Original, Bottom: Result of instance segmentation framework) (A-B) False negatives resulting from sudden movements of a surgical instrument. (C) False positive, in which part of a surgical instrument is recognized as a single object. (D) False positive case, in which a non-surgical instrument is recognized as a surgical instrument.

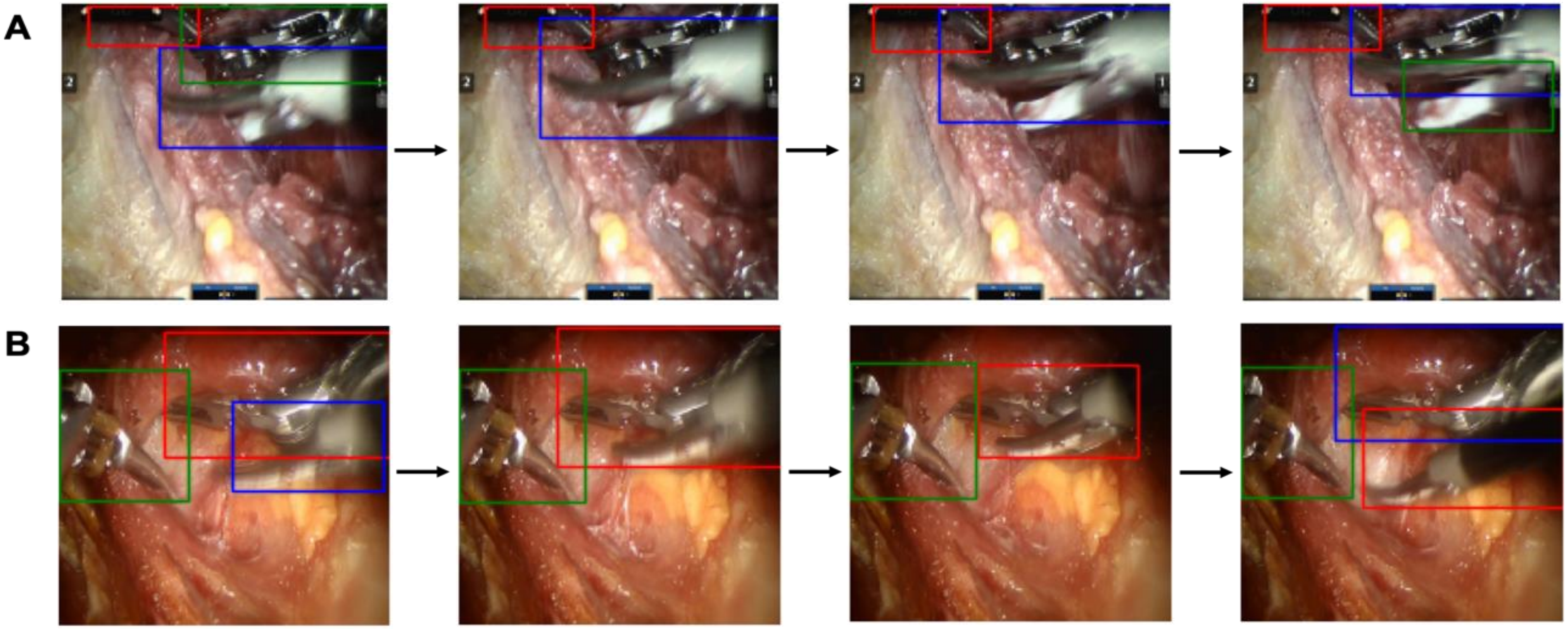


Figure 3.18 Errors in the tracking framework. (A-B) Errors in which the identity of surgical instruments was switched following occlusion.

3.5 Conclusion

The proposed system can track the surgical instruments using deep learning-based visual tracking methods and enable the automatic and quantitative assessment of robotic surgical skill. It is expected that the proposed system will effectively educate students who need robotic surgery training, and will improve surgical skill of surgeons during the performance of robotic surgery.

* Large sections of this chapter were published previously in *Journal of Clinical Medicine*. (Dongheon Lee, Hyeong Won Yu, et al. “Evaluation of Surgical Skills during Robotic Surgery by Deep Learning-Based Multiple Surgical Instrument Tracking in Training and Actual Operations”, *Journal of Clinical Medicine*, 2020, 9, 1964) [138]

Chapter 4 Summary and Future Works

4.1 Thesis Summary

The goal of this study was to develop deep learning methods to improve the performance of clinicians and to evaluate the effect of the proposed methods on the outcome. In the first study, we developed an algorithm to classify the type of colorectal polyp images taken by a narrow-band imaging colonoscopy. The developed method not only showed improved performance as compared to results seen in previous studies, but also aided in endoscopists' diagnosis process by presenting a visually basis of the AI prediction. The effectiveness of the method was verified through clinical evaluations, which showed that the average diagnostic accuracy of the endoscopists was improved and that the average diagnosis time was shortened as well.

In the second study, we developed a deep learning-based surgical instruments tracking algorithm in robotic surgery videos. Since the motion of the surgical instrument is significantly related to the surgeon's skill, it was possible to develop an evaluation model by calculating the motion metrics. As a result, quantitative and automatic evaluation model for surgeons' robotic surgical skill was demonstrated based on the tracking algorithm. The proposed system may pose as a direction towards quantitative evaluation of clinicians, thus improving the robotic surgical field as a whole.

In conclusion, this study proved that the proposed deep learning methods can be effective tools for training and improving the performance of clinicians' skills, and also showed the possibility of replacing previous methods used in clinical fields by presenting novel concepts and verifying improved performances.

4.2 Limitations and Future works

In order to apply deep learning methods for medical images analysis in clinical practice including colonoscopic images and robotic surgery videos, several limitations should be addressed. The first is the set of issues particular to medical image datasets. For developing and verifying deep learning-based algorithms, large-scale and high-quality images as well as data distribution generalizable to new populations [139-142] are required. It is also necessary to collect datasets from multiple sources [139], and to transcend nationality, gender and race [143]. In addition, in the development process, objects in the background including noise that may impede the training, must be removed in advance [144, 145]. Furthermore, due to the nature of medical images, and specially of endoscopic images, it is necessary to apply novel approaches in conjunction with traditional deep learning methods, because issues exists such as the uncertainty of cell stage during differentiation [86, 146].

Second, deep learning methods must be interpretable. Since traditional deep learning approaches regard the neural network as a black box despite its high performance [81], the process of prediction must be interpreted such that clinicians can understand the basis of the prediction result [58, 82].

Next is the liability issue [147-149]. The question of who takes responsibility for the medical treatment based on the prediction results by AI is still on debate. Responsibility can be held by doctor, AI systems, or insurance companies, and the related issue is expected to be established in the near future [150].

Finally, achieving robust regulation and rigorous quality control are necessary [151]. AI systems need to be systematically managed, such that datasets can be continuous, periodical, and system-widely updated, as being attempted by the U.S. Food and Drug Administration. Through such control, AI systems can be generalized and overcome biases, thereby enabling higher performance and stable operation [152].

Bibliography

[1] J.-G. Lee *et al.*, "Deep learning in medical imaging: general overview," *Korean journal of radiology,* vol. 18, no. 4, pp. 570-584, 2017.

[2] Y. Bengio and Y. LeCun, "Scaling learning algorithms towards AI," *Large-scale kernel machines,* vol. 34, no. 5, pp. 1-41, 2007.

[3] H. Takiyama *et al.*, "Automatic anatomical classification of esophagogastroduodenoscopy images using deep convolutional neural networks," *Scientific reports,* vol. 8, no. 1, pp. 1-8, 2018.

[4] G. Litjens *et al.*, "A survey on deep learning in medical image analysis," *Medical image analysis,* vol. 42, pp. 60-88, 2017.

[5] A. Fourcade and R. Khonsari, "Deep learning in medical image analysis: A third eye for doctors," *Journal of stomatology, oral and maxillofacial surgery,* vol. 120, no. 4, pp. 279-288, 2019.

[6] A. Maier, C. Syben, T. Lasser, and C. Riess, "A gentle introduction to deep learning in medical image processing," *Zeitschrift für Medizinische Physik,* vol. 29, no. 2, pp. 86-101, 2019.

[7] V. Gulshan *et al.*, "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," *Jama,* vol. 316, no. 22, pp. 2402-2410, 2016.

[8] A. Esteva *et al.*, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature,* vol. 542, no. 7639, pp. 115-118, 2017.

[9] X. Yi, E. Walia, and P. Babyn, "Generative adversarial network in medical imaging: A review," *Medical image analysis,* vol. 58, p. 101552, 2019.

[10] P. Brandao *et al.*, "Fully convolutional neural networks for polyp segmentation in colonoscopy," in *Medical Imaging 2017: Computer-Aided Diagnosis*, 2017, vol. 10134, p. 101340F: International Society for Optics and Photonics.

[11] M. F. Byrne *et al.*, "Real-time differentiation of adenomatous and hyperplastic diminutive colorectal polyps during analysis of unaltered

videos of standard colonoscopy using a deep learning model," *Gut,* vol. 68, no. 1, pp. 94-100, 2019.

[12] Y. Komeda *et al.*, "Computer-aided diagnosis based on convolutional neural network system for colorectal polyp classification: preliminary experience," *Oncology,* vol. 93, no. Suppl. 1, pp. 30-34, 2017.

[13] S. Fan, L. Xu, Y. Fan, K. Wei, and L. Li, "Computer-aided detection of small intestinal ulcer and erosion in wireless capsule endoscopy images," *Physics in Medicine & Biology,* vol. 63, no. 16, p. 165001, 2018.

[14] R. Zhang *et al.*, "Automatic detection and classification of colorectal polyps by transferring low-level CNN features from nonmedical domain," *IEEE journal of biomedical and health informatics,* vol. 21, no. 1, pp. 41-47, 2016.

[15] L. Yu, H. Chen, Q. Dou, J. Qin, and P. A. Heng, "Integrating online and offline three-dimensional deep learning for automated polyp detection in colonoscopy videos," *IEEE journal of biomedical and health informatics,* vol. 21, no. 1, pp. 65-75, 2016.

[16] I. Boonpogmanee, "Fully Convolutional Neural Networks for Semantic Segmentation of Polyp Images Taken During Colonoscopy: 2760," *American Journal of Gastroenterology,* vol. 113, p. S1532, 2018.

[17] R. F. Mansour, "Deep-learning-based automatic computer-aided diagnosis system for diabetic retinopathy," *Biomedical engineering letters,* vol. 8, no. 1, pp. 41-57, 2018.

[18] A. Rau *et al.*, "Implicit domain adaptation with conditional generative adversarial networks for depth prediction in endoscopy," *International journal of computer assisted radiology and surgery,* vol. 14, no. 7, pp. 1167-1176, 2019.

[19] J. Choi *et al.*, "Convolutional Neural Network Technology in Endoscopic Imaging: Artificial Intelligence for Endoscopy," *Clinical Endoscopy,* vol. 53, no. 2, p. 117, 2020.

[20] E. J. Topol, "High-performance medicine: the convergence of human and artificial intelligence," *Nature medicine,* vol. 25, no. 1, pp. 44-56, 2019.

[21] A. P. Twinanda, S. Shehata, D. Mutter, J. Marescaux, M. De Mathelin, and N. Padoy, "Endonet: A deep architecture for recognition tasks on laparoscopic videos," *IEEE transactions on medical imaging,* vol. 36, no. 1, pp. 86-97, 2017.

[22] D. Sarikaya, J. J. Corso, and K. A. Guru, "Detection and localization of robotic tools in robot-assisted surgery videos using deep neural networks for region proposal and detection," *IEEE transactions on medical imaging,* vol. 36, no. 7, pp. 1542-1549, 2017.

[23] A. Shvets, A. Rakhlin, A. A. Kalinin, and V. Iglovikov, "Automatic Instrument Segmentation in Robot-Assisted Surgery Using Deep Learning," *arXiv preprint arXiv:1803.01207,* 2018.

[24] H. Law, K. Ghani, and J. Deng, "Surgeon technical skill assessment using computer vision based analysis," in *Machine learning for healthcare conference*, 2017, pp. 88-99.

[25] A. Reiter, P. K. Allen, and T. Zhao, "Articulated surgical tool detection using virtually-rendered templates," in *Computer Assisted Radiology and Surgery (CARS)*, 2012, pp. 1-8.

[26] H. I. Fawaz, G. Forestier, J. Weber, L. Idoumghar, and P.-A. Muller, "Evaluating surgical skills from kinematic data using convolutional neural networks," *arXiv preprint arXiv:1806.02750,* 2018.

[27] H. C. Lin, I. Shafran, D. Yuh, and G. D. Hager, "Towards automatic skill evaluation: Detection and segmentation of robot-assisted surgical motions," *Computer Aided Surgery,* vol. 11, no. 5, pp. 220-230, 2006.

[28] R. Kumar *et al.*, "Objective measures for longitudinal assessment of robotic surgery training," *The Journal of thoracic and cardiovascular surgery,* vol. 143, no. 3, pp. 528-534, 2012.

[29] A. J. Hung *et al.*, "Experts vs super-experts: differences in automated performance metrics and clinical outcomes for robot-assisted radical prostatectomy," *BJU international,* vol. 123, no. 5, pp. 861-868, 2019.

[30] K. Mishra, R. Sathish, and D. Sheet, "Learning latent temporal connectionism of deep residual visual abstractions for identifying surgical tools in laparoscopy procedures," in *Proceedings of the IEEE*

Conference on Computer Vision and Pattern Recognition Workshops, 2017, pp. 58-65.

[31] M. Sahu, A. Mukhopadhyay, A. Szengel, and S. Zachow, "Addressing multi-label imbalance problem of surgical tool detection using CNN," *International journal of computer assisted radiology and surgery,* vol. 12, no. 6, pp. 1013-1020, 2017.

[32] B. Choi, K. Jo, S. Choi, and J. Choi, "Surgical-tools detection based on Convolutional Neural Network in laparoscopic robot-assisted surgery," in *Engineering in Medicine and Biology Society (EMBC), 2017 39th Annual International Conference of the IEEE*, 2017, pp. 1756-1759: IEEE.

[33] L. C. García-Peraza-Herrera *et al.*, "Real-time segmentation of non-rigid surgical tools based on deep learning and tracking," in *International Workshop on Computer-Assisted and Robotic Endoscopy*, 2016, pp. 84-95: Springer.

[34] T. Kurmann *et al.*, "Simultaneous recognition and pose estimation of instruments in minimally invasive surgery," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2017, pp. 505-513: Springer.

[35] F. Yu *et al.*, "Assessment of automated identification of phases in videos of cataract surgery using machine learning and deep learning techniques," *JAMA network open,* vol. 2, no. 4, pp. e191860-e191860, 2019.

[36] S. Khalid, M. Goldenberg, T. Grantcharov, B. Taati, and F. Rudzicz, "Evaluation of Deep Learning Models for Identifying Surgical Actions and Measuring Performance," *JAMA Network Open,* vol. 3, no. 3, pp. e201664-e201664, 2020.

[37] J. Martin *et al.*, "Objective structured assessment of technical skill (OSATS) for surgical residents," *British journal of surgery,* vol. 84, no. 2, pp. 273-278, 1997.

[38] A. C. Goh, D. W. Goldfarb, J. C. Sander, B. J. Miles, and B. J. Dunkin, "Global evaluative assessment of robotic skills: validation of a clinical

assessment tool to measure robotic surgical skills," *The Journal of urology,* vol. 187, no. 1, pp. 247-252, 2012.

[39] K. A. Cronin *et al.*, "Annual Report to the Nation on the Status of Cancer, part I: National cancer statistics," *Cancer,* vol. 124, no. 13, pp. 2785-2800, 2018.

[40] J. J. Sung, J. Y. Lau, K. Goh, W. Leung, and A. P. W. G. o. C. Cancer, "Increasing incidence of colorectal cancer in Asia: implications for screening," *The lancet oncology,* vol. 6, no. 11, pp. 871-876, 2005.

[41] A. Leslie, F. Carey, N. Pratt, and R. Steele, "The colorectal adenoma–carcinoma sequence," *British Journal of Surgery,* vol. 89, no. 7, pp. 845-860, 2002.

[42] D. A. Corley *et al.*, "Adenoma detection rate and risk of colorectal cancer and death," *New england journal of medicine,* vol. 370, no. 14, pp. 1298-1306, 2014.

[43] D. Lieberman, M. Moravec, J. Holub, L. Michaels, and G. Eisen, "Polyp size and advanced histology in patients undergoing colonoscopy screening: implications for CT colonography," *Gastroenterology,* vol. 135, no. 4, pp. 1100-1105, 2008.

[44] P. L. Ponugoti, O. W. Cummings, and D. K. Rex, "Risk of cancer in small and diminutive colorectal polyps," *Digestive and Liver Disease,* vol. 49, no. 1, pp. 34-37, 2017.

[45] D. K. Rex, "Narrow-band imaging without optical magnification for histologic analysis of colorectal polyps," *Gastroenterology,* vol. 136, no. 4, pp. 1174-1181, 2009.

[46] P. Ponugoti *et al.*, "Disagreement between high confidence endoscopic adenoma prediction and histopathological diagnosis in colonic lesions≤ 3 mm in size," *Endoscopy,* vol. 51, no. 03, pp. 221-226, 2019.

[47] N. Shahidi, D. K. Rex, T. Kaltenbach, A. Rastogi, S. H. Ghalehjegh, and M. F. Byrne, "Use of Endoscopic Impression, Artificial Intelligence, and Pathologist Interpretation to Resolve Discrepancies Between Endoscopy and Pathology Analyses of Diminutive Colorectal Polyps," *Gastroenterology,* vol. 158, no. 3, pp. 783-785. e1, 2020.

[48] D. G. Hewett *et al.*, "Validation of a simple classification system for endoscopic diagnosis of small colorectal polyps using narrow-band imaging," *Gastroenterology,* vol. 143, no. 3, pp. 599-607. e1, 2012.

[49] K. Gono, "Narrow band imaging: technology basis and research and development history," *Clinical endoscopy,* vol. 48, no. 6, p. 476, 2015.

[50] A. Ignjatovic *et al.*, "Development and validation of a training module on the use of narrow-band imaging in differentiation of small adenomas from hyperplastic colorectal polyps," *Gastrointestinal endoscopy,* vol. 73, no. 1, pp. 128-133, 2011.

[51] J. L. Vleugels *et al.*, "Effects of training and feedback on accuracy of predicting rectosigmoid neoplastic lesions and selection of surveillance intervals by endoscopists performing optical diagnosis of diminutive polyps," *Gastroenterology,* vol. 154, no. 6, pp. 1682-1693. e1, 2018.

[52] M. Misawa *et al.*, "Characterization of colorectal lesions using a computer-aided diagnostic system for narrow-band imaging endocytoscopy," *Gastroenterology,* vol. 150, no. 7, pp. 1531-1532. e3, 2016.

[53] A. Esteva *et al.*, "A guide to deep learning in healthcare," *Nature medicine,* vol. 25, no. 1, pp. 24-29, 2019.

[54] P.-J. Chen, M.-C. Lin, M.-J. Lai, J.-C. Lin, H. H.-S. Lu, and V. S. Tseng, "Accurate classification of diminutive colorectal polyps using computer-aided analysis," *Gastroenterology,* vol. 154, no. 3, pp. 568-575, 2018.

[55] Y. Mori *et al.*, "Impact of an automated system for endocytoscopic diagnosis of small colorectal lesions: an international web-based study," *Endoscopy,* vol. 48, no. 12, pp. 1110-1118, 2016.

[56] Y. Mori, S.-e. Kudo, T. M. Berzin, M. Misawa, and K. Takeda, "Computer-aided diagnosis for colonoscopy," *Endoscopy,* vol. 49, no. 08, pp. 813-819, 2017.

[57] D. Shen, G. Wu, and H.-I. Suk, "Deep learning in medical image analysis," *Annual review of biomedical engineering,* vol. 19, pp. 221-248, 2017.

[58] R. Sayres *et al.*, "Using a deep learning algorithm and integrated gradients explanation to assist grading for diabetic retinopathy," *Ophthalmology,* vol. 126, no. 4, pp. 552-564, 2019.

[59] M. Billah and S. Waheed, "Gastrointestinal polyp detection in endoscopic images using an improved feature extraction method," *Biomedical engineering letters,* vol. 8, no. 1, pp. 69-75, 2018.

[60] P. Wang *et al.*, "Development and validation of a deep-learning algorithm for the detection of polyps during colonoscopy," *Nature biomedical engineering,* vol. 2, no. 10, p. 741, 2018.

[61] J. H. Bae *et al.*, "Improved real-time optical diagnosis of colorectal polyps following a comprehensive training program," *Clinical Gastroenterology and Hepatology,* vol. 17, no. 12, pp. 2479-2488. e4, 2019.

[62] E. Castro, J. S. Cardoso, and J. C. Pereira, "Elastic deformations for data augmentation in breast cancer mass detection," in *2018 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*, 2018, pp. 230-234: IEEE.

[63] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2818-2826.

[64] H. Pham, M. Y. Guan, B. Zoph, Q. V. Le, and J. Dean, "Efficient neural architecture search via parameter sharing," *arXiv preprint arXiv:1802.03268,* 2018.

[65] W. Zaremba, I. Sutskever, and O. Vinyals, "Recurrent neural network regularization," *arXiv preprint arXiv:1409.2329,* 2014.

[66] R. S. Sutton and A. G. Barto, *Introduction to reinforcement learning* (no. 4). MIT press Cambridge, 1998.

[67] B. Zoph, V. Vasudevan, J. Shlens, and Q. V. Le, "Learning transferable architectures for scalable image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 8697-8710.

[68] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980,* 2014.

[69] Y. S. Aurelio, G. M. de Almeida, C. L. de Castro, and A. P. Braga, "Learning from imbalanced data sets with weighted cross-entropy function," *Neural Processing Letters,* pp. 1-13, 2019.

[70] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 618-626.

[71] L. v. d. Maaten and G. Hinton, "Visualizing data using t-SNE," *Journal of machine learning research,* vol. 9, no. Nov, pp. 2579-2605, 2008.

[72] S.-W. L. Ung Lee, Young-Chul Shin, Dong-Won Shin, Kang Seob Oh, Sun-Young Kim, Young Hwan Kim, Sang Won Jeon "Reliability and Validity of Korean version of GRIT," *Anxiety and Mood,* vol. 15, no. 1, pp. 53-60, 2019.

[73] L. Halliday, A. Walker, S. Vig, J. Hines, and J. Brecknell, "Grit and burnout in UK doctors: a cross-sectional study across specialties and stages of training," *Postgraduate medical journal,* vol. 93, no. 1101, pp. 389-394, 2017.

[74] A. L. Duckworth, C. Peterson, M. D. Matthews, and D. R. Kelly, "Grit: perseverance and passion for long-term goals," *Journal of personality and social psychology,* vol. 92, no. 6, p. 1087, 2007.

[75] L. R. Miller-Matero, S. Martinez, L. MacLean, K. Yaremchuk, and A. B. Ko, "Grit: A predictor of medical student performance," *Education for Health,* vol. 31, no. 2, p. 109, 2018.

[76] A. Salles *et al.*, "Grit as a predictor of risk of attrition in surgical residency," *The American Journal of Surgery,* vol. 213, no. 2, pp. 288-291, 2017.

[77] A. Dam, T. Perera, M. Jones, M. Haughy, and T. Gaeta, "The Relationship Between Grit, Burnout, and Well-being in Emergency Medicine Residents," *AEM education and training,* vol. 3, no. 1, pp. 14-19, 2019.

[78] A. L. Duckworth, C. Peterson, M. D. Matthews, and D. R. Kelly, "Grit: perseverance and passion for long-term goals," *J Pers Soc Psychol,* vol. 92, no. 6, pp. 1087-101, Jun 2007.

[79] J. Wang and L. Perez, "The effectiveness of data augmentation in image classification using deep learning," *Convolutional Neural Networks Vis. Recognit,* 2017.

[80] G. Urban *et al.*, "Deep learning localizes and identifies polyps in real time with 96% accuracy in screening colonoscopy," *Gastroenterology,* vol. 155, no. 4, pp. 1069-1078. e8, 2018.

[81] D. Castelvecchi, "Can we open the black box of AI?," *Nature News,* vol. 538, no. 7623, p. 20, 2016.

[82] R. Poplin *et al.*, "Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning," *Nature Biomedical Engineering,* vol. 2, no. 3, p. 158, 2018.

[83] T. Kuiper *et al.*, "Accuracy for optical diagnosis of small colorectal polyps in nonacademic settings," *Clinical Gastroenterology and Hepatology,* vol. 10, no. 9, pp. 1016-1020, 2012.

[84] J. N. Rogart *et al.*, "Narrow-band imaging without high magnification to differentiate polyps during real-time colonoscopy: improvement with experience," *Gastrointestinal endoscopy,* vol. 68, no. 6, pp. 1136-1145, 2008.

[85] J. Adebayo, J. Gilmer, M. Muelly, I. Goodfellow, M. Hardt, and B. Kim, "Sanity checks for saliency maps," in *Advances in Neural Information Processing Systems*, 2018, pp. 9505-9515.

[86] E. Begoli, T. Bhattacharya, and D. Kusnezov, "The need for uncertainty quantification in machine-assisted medical decision making," *Nature Machine Intelligence,* vol. 1, no. 1, pp. 20-23, 2019.

[87] J. Wagner, J. M. Kohler, T. Gindele, L. Hetzel, J. T. Wiedemer, and S. Behnke, "Interpretable and Fine-Grained Visual Explanations for Convolutional Neural Networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 9097-9107.

[88] A. Kendall and Y. Gal, "What uncertainties do we need in bayesian deep learning for computer vision?," in *Advances in neural information processing systems*, 2017, pp. 5574-5584.

[89] E. H. Jin *et al.*, "Improved accuracy in optical diagnosis of colorectal polyps using convolutional neural networks with visual explanations," *Gastroenterology,* 2020.

[90] T. L. Ghezzi and O. C. Corleta, "30 years of robotic surgery," *World journal of surgery,* vol. 40, no. 10, pp. 2550-2557, 2016.

[91] A. Ng and P. Tam, "Current status of robot-assisted surgery," *Hong Kong Med J,* vol. 20, no. 3, pp. 241-250, 2014.

[92] L. I. Pernar, F. C. Robertson, A. Tavakkoli, E. G. Sheu, D. C. Brooks, and D. S. Smink, "An appraisal of the learning curve in robotic general surgery," *Surgical endoscopy,* vol. 31, no. 11, pp. 4583-4596, 2017.

[93] C. Perrenot *et al.*, "The virtual reality simulator dV-Trainer® is a valid assessment tool for robotic surgical skills," *Surgical endoscopy,* vol. 26, no. 9, pp. 2587-2593, 2012.

[94] W. M. Brinkman, J.-M. Luursema, B. Kengen, B. M. Schout, J. A. Witjes, and R. L. Bekkers, "da Vinci skills simulator for assessing learning curve and criterion-based training of robotic basic skills," *Urology,* vol. 81, no. 3, pp. 562-566, 2013.

[95] Z. Hilal, A. K. Kumpernatz, G. A. Rezniczek, C. Cetin, E.-K. Tempfer-Bentz, and C. B. Tempfer, "A randomized comparison of video demonstration versus hands-on training of medical students for vacuum delivery using Objective Structured Assessment of Technical Skills (OSATS)," *Medicine,* vol. 96, no. 11, 2017.

[96] S. L. McGregor, *Understanding and evaluating research: A critical guide*. SAGE Publications, 2017.

[97] J. R. Mark, D. C. Kelly, E. J. Trabulsi, P. J. Shenot, and C. D. Lallas, "The effects of fatigue on robotic surgical skill training in Urology residents," *Journal of robotic surgery,* vol. 8, no. 3, pp. 269-275, 2014.

[98] S.-K. Jun *et al.*, "Robotic minimally invasive surgical skill assessment based on automated video-analysis motion studies," in *2012 4th IEEE*

RAS & EMBS International Conference on Biomedical Robotics and Biomechatronics (BioRob), 2012, pp. 25-31: IEEE.

[99] S. Speidel, M. Delles, C. Gutt, and R. Dillmann, "Tracking of instruments in minimally invasive surgery for surgical skill analysis," in *International Workshop on Medical Imaging and Virtual Reality*, 2006, pp. 148-155: Springer.

[100] J. Ryu, J. Choi, and H. C. Kim, "Endoscopic vision-based tracking of multiple surgical instruments during robot-assisted surgery," *Artificial organs,* vol. 37, no. 1, pp. 107-112, 2013.

[101] L. C. García-Peraza-Herrera *et al.*, "ToolNet: holistically-nested real-time segmentation of robotic surgical tools," in *Intelligent Robots and Systems (IROS), 2017 IEEE/RSJ International Conference on*, 2017, pp. 5717-5722: IEEE.

[102] D. Pakhomov, V. Premachandran, M. Allan, M. Azizian, and N. Navab, "Deep residual learning for instrument segmentation in robotic surgery," *arXiv preprint arXiv:1703.08580,* 2017.

[103] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, and Q. Tian, "Scalable person re-identification: A benchmark," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1116-1124.

[104] N. Wojke, A. Bewley, and D. Paulus, "Simple online and realtime tracking with a deep association metric," in *Image Processing (ICIP), 2017 IEEE International Conference on*, 2017, pp. 3645-3649: IEEE.

[105] H. W. Yu *et al.*, "Development of a surgical training model for bilateral axillo-breast approach robotic thyroidectomy," *Surgical endoscopy,* vol. 32, no. 3, pp. 1360-1367, 2018.

[106] K. E. Lee, E. Kim, D. H. Koo, J. Y. Choi, K. H. Kim, and Y.-K. Youn, "Robotic thyroidectomy by bilateral axillo-breast approach: review of 1026 cases and surgical completeness," *Surgical endoscopy,* vol. 27, no. 8, pp. 2955-2962, 2013.

[107] I. Oropesa *et al.*, "EVA: laparoscopic instrument tracking based on endoscopic video analysis for psychomotor skills assessment," *Surgical endoscopy,* vol. 27, no. 3, pp. 1029-1039, 2013.

[108] A. Jin *et al.*, "Tool Detection and Operative Skill Assessment in Surgical Videos Using Region-Based Convolutional Neural Networks," *arXiv preprint arXiv:1802.08774,* 2018.

[109] S. Y.-W. Liu and J. S. Kim, "Bilateral axillo-breast approach robotic thyroidectomy: review of evidences," *Gland surgery,* vol. 6, no. 3, p. 250, 2017.

[110] Q. He *et al.*, "Robotic lateral cervical lymph node dissection via bilateral axillo-breast approach for papillary thyroid carcinoma: a single-center experience of 260 cases," *Journal of Robotic Surgery,* pp. 1-7, 2019.

[111] N. Christou and M. Mathonnet, "Complications after total thyroidectomy," *Journal of visceral surgery,* vol. 150, no. 4, pp. 249-256, 2013.

[112] M. Allan *et al.*, "2017 Robotic instrument segmentation challenge," *arXiv preprint arXiv:1902.06426,* 2019.

[113] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *Computer Vision (ICCV), 2017 IEEE International Conference on*, 2017, pp. 2980-2988: IEEE.

[114] G. Wang, J. Lai, P. Huang, and X. Xie, "Spatial-temporal person re-identification," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2019, vol. 33, pp. 8933-8940.

[115] I. Laina *et al.*, "Concurrent segmentation and localization for tracking of surgical instruments," in *International conference on medical image computing and computer-assisted intervention*, 2017, pp. 664-672: Springer.

[116] R. Girshick, "Fast r-cnn," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1440-1448.

[117] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770-778.

[118] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proceedings of the*

IEEE conference on computer vision and pattern recognition, 2017, pp. 2117-2125.

[119] G. Welch and G. Bishop, "An introduction to the Kalman filter," 1995.

[120] H. W. Kuhn, "The Hungarian method for the assignment problem," *Naval research logistics quarterly,* vol. 2, no. 1-2, pp. 83-97, 1955.

[121] X. Peng, L. Wang, X. Wang, and Y. Qiao, "Bag of visual words and fusion methods for action recognition: Comprehensive study and good practice," *Computer Vision and Image Understanding,* vol. 150, pp. 109-125, 2016.

[122] E. Rublee, V. Rabaud, K. Konolige, and G. R. Bradski, "ORB: An efficient alternative to SIFT or SURF," in *ICCV*, 2011, vol. 11, no. 1, p. 2: Citeseer.

[123] P. K. Saha, G. Borgefors, and G. S. di Baja, "A survey on skeletonization algorithms and their applications," *Pattern Recognition Letters,* vol. 76, pp. 3-12, 2016.

[124] J.-C. Yoo and T. H. Han, "Fast normalized cross-correlation," *Circuits, systems and signal processing,* vol. 28, no. 6, p. 819, 2009.

[125] C. Yu *et al.*, "Acral melanoma detection using a convolutional neural network for dermoscopy images," *PloS one,* vol. 13, no. 3, p. e0193321, 2018.

[126] Y. Yamazaki *et al.*, "Automated Surgical Instrument Detection from Laparoscopic Gastrectomy Video Images Using an Open Source Convolutional Neural Network Platform," *Journal of the American College of Surgeons,* 2020.

[127] S. L. Vernez, V. Huynh, K. Osann, Z. Okhunov, J. Landman, and R. V. Clayman, "C-SATS: assessing surgical skills among urology residency applicants," *Journal of endourology,* vol. 31, no. S1, pp. S-95-S-100, 2017.

[128] M. G. Goldenberg *et al.*, "Feasibility of expert and crowd-sourced review of intraoperative video for quality improvement of intracorporeal urinary diversion during robotic radical cystectomy," *Canadian Urological Association Journal,* vol. 11, no. 10, p. 331, 2017.

[129] J. B. Pagador *et al.*, "Decomposition and analysis of laparoscopic suturing task using tool-motion analysis (TMA): improving the objective assessment," *International journal of computer assisted radiology and surgery,* vol. 7, no. 2, pp. 305-313, 2012.

[130] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," *Journal of artificial intelligence research,* vol. 16, pp. 321-357, 2002.

[131] S. Paisitkriangkrai, C. Shen, and A. Van Den Hengel, "Learning to rank in person re-identification with metric ensembles," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 1846-1855.

[132] D. Lee *et al.*, "Vision-based tracking system for augmented reality to localize recurrent laryngeal nerve during robotic thyroid surgery," *Scientific Reports,* vol. 10, no. 1, pp. 1-7, 2020.

[133] A. Reiter, P. K. Allen, and T. Zhao, "Appearance learning for 3D tracking of robotic surgical tools," *The International Journal of Robotics Research,* vol. 33, no. 2, pp. 342-356, 2014.

[134] A. Reiter, P. K. Allen, and T. Zhao, "Learning features on robotic surgical tools," in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society Conference on*, 2012, pp. 38-43: IEEE.

[135] I. Nisky, M. H. Hsieh, and A. M. Okamura, "The effect of a robot-assisted surgical system on the kinematics of user movements," in *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2013, pp. 6257-6260: IEEE.

[136] M. Allan, S. Ourselin, S. Thompson, D. J. Hawkes, J. Kelly, and D. Stoyanov, "Toward detection and localization of instruments in minimally invasive surgery," *IEEE Transactions on Biomedical Engineering,* vol. 60, no. 4, pp. 1050-1058, 2013.

[137] M. Allan *et al.*, "Image based surgical instrument pose estimation with multi-class labelling and optical flow," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2015, pp. 331-338: Springer.

[138] D. Lee, H. W. Yu, H. Kwon, H.-J. Kong, K. E. Lee, and H. C. Kim, "Evaluation of Surgical Skills during Robotic Surgery by Deep Learning-Based Multiple Surgical Instrument Tracking in Training and Actual Operations," *Journal of Clinical Medicine,* vol. 9, no. 6, p. 1964, 2020.

[139] S. Chilamkurthy *et al.*, "Deep learning algorithms for detection of critical findings in head CT scans: a retrospective study," *The Lancet,* vol. 392, no. 10162, pp. 2388-2396, 2018.

[140] T. P. Debray, Y. Vergouwe, H. Koffijberg, D. Nieboer, E. W. Steyerberg, and K. G. Moons, "A new framework to enhance the interpretation of external validation studies of clinical prediction models," *Journal of clinical epidemiology,* vol. 68, no. 3, pp. 279-289, 2015.

[141] E. J. Hwang *et al.*, "Development and validation of a deep learning–based automated detection algorithm for major thoracic diseases on chest radiographs," *JAMA network open,* vol. 2, no. 3, pp. e191095-e191095, 2019.

[142] D. W. Kim, H. Y. Jang, K. W. Kim, Y. Shin, and S. H. Park, "Design characteristics of studies reporting the performance of artificial intelligence algorithms for diagnostic analysis of medical images: results from recently published papers," *Korean journal of radiology,* vol. 20, no. 3, pp. 405-410, 2019.

[143] S. Barocas and A. D. Selbst, "Big data's disparate impact," *Calif. L. Rev.,* vol. 104, p. 671, 2016.

[144] M. T. Ribeiro, S. Singh, and C. Guestrin, "" Why should i trust you?" Explaining the predictions of any classifier," in *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, 2016, pp. 1135-1144.

[145] J. K. Winkler *et al.*, "Association between surgical skin markings in dermoscopic images and diagnostic performance of a deep learning convolutional neural network for melanoma recognition," *JAMA dermatology,* vol. 155, no. 10, pp. 1135-1141, 2019.

[146] G. Carneiro, L. Z. C. T. Pu, R. Singh, and A. Burt, "Deep Learning Uncertainty and Confidence Calibration for the Five-class Polyp

Classification from Colonoscopy," *Medical Image Analysis,* p. 101653, 2020.

[147] M. Coeckelbergh, "Artificial Intelligence, Responsibility Attribution, and a Relational Justification of Explainability," *Science and engineering ethics,* pp. 1-18, 2019.

[148] E. Neri, F. Coppola, V. Miele, C. Bibbolino, and R. Grassi, "Artificial intelligence: Who is responsible for the diagnosis?," ed: Springer, 2020.

[149] W. Orr and J. L. Davis, "Attributions of ethical responsibility by Artificial Intelligence practitioners," *Information, Communication & Society,* pp. 1-17, 2020.

[150] W. N. Price, S. Gerke, and I. G. Cohen, "Potential liability for physicians using artificial intelligence," *Jama,* vol. 322, no. 18, pp. 1765-1766, 2019.

[151] U. FDA, "Proposed regulatory framework for modifications to artificial intelligence/machine learning (AI/ML)-based software as a medical device (SaMD)," ed: FDA, 2019.

[152] I. Chen, F. D. Johansson, and D. Sontag, "Why is my classifier discriminatory?," in *Advances in Neural Information Processing Systems*, 2018, pp. 3539-3550.

초 록

본 논문은 의료진의 임상술기 능력을 향상시키기 위하여 새로운 딥러닝 기법들을 제안하고 다음 두 가지 실례에 대해 적용하여 그 결과를 평가하였다.

첫 번째 연구에서는 대장내시경으로 광학 진단 시, 내시경 전문의의 진단 능력을 향상시키기 위하여 딥러닝 기반의 용종 분류 알고리즘을 개발하고, 내시경 전문의의 진단 능력 향상 여부를 검증하고자 하였다. 대장내시경 검사로 암종으로 증식할 수 있는 선종과 과증식성 용종을 진단하는 것은 중요하다. 본 연구에서는 협대역 영상 내시경으로 촬영한 대장 용종 영상으로 합성곱 신경망을 학습하여 분류 알고리즘을 개발하였다. 제안하는 알고리즘은 자동 기계학습 (AutoML) 방법으로, 대장 용종 영상에 최적화된 합성곱 신경망 구조를 찾고 신경망의 가중치를 학습하였다. 또한 기울기-가중치 클래스 활성화 맵핑 기법을 이용하여 개발한 합성곱 신경망 결과의 확률적 근거를 용종 위치에 시각적으로 나타나도록 함으로 내시경 전문의의 진단을 돕도록 하였다. 마지막으로, 숙련도 그룹별로 내시경 전문의가 용종 분류 알고리즘의 결과를 참고하였을 때 진단 능력이 향상되었는지 비교 실험을 진행하였고, 모든 그룹에서 유의미하게 진단 정확도가 향상되고 진단 시간이 단축되었음을 확인하였다.

두 번째 연구에서는 로봇수술 동영상에서 수술도구 위치 추적 알고리즘을 개발하고, 획득한 수술도구의 움직임 정보를 바탕으로 수술자의 숙련도를 정량적으로 평가하는 모델을 제안하였다. 수술도구의 움직임은 수술자의 로봇수술 숙련도를 평가하기 위한 주요한 정보이다. 따라서 본 연구는 딥러닝 기반의 자동 수술도구 추적 알고리즘을 개발하였으며, 다음 두가지 선행연구의 한계점을 극복하였다. 인스턴스 분할 (Instance Segmentation) 프레임웍을 개발하여 폐색 (Occlusion) 문제를 해결하였고, 추적기 (Tracker)와 재식별화 (Re-Identification) 알고리즘으로 구성된 추적 프레임웍을 개발하여 동영상에서 추적하는 수술도구의 종류가 유지되도록 하였다. 또한 로봇수술 동영상의 특수성을 고려하여 수술도구의 움직임을 획득하기위해 수술도구 끝 위치와 로봇 팔-인디케이터 (Arm-Indicator) 인식 알고리즘을 개발하였다. 제안하는 알고리즘의 성능은 예측한 수술도구 끝 위치와 정답 위치 간의 평균 제곱근 오차, 곡선 아래 면적, 피어슨 상관분석으로 평가하였다. 마지막으로, 수술도구의 움직임으로부터 움직임 지표를 계산하고 이를 바탕으로 기계학습 기반의 로봇수술 숙련도 평가 모델을 개발하였다. 개발한 평가 모델은 기존의 Objective Structured Assessment of Technical Skill (OSATS), Global Evaluative Assessment of Robotic Surgery (GEARS) 평가 방법과 유사한 성능을 보임을 확인하였다.

본 논문은 의료진의 임상술기 능력을 향상시키기 위하여 대장 용종 영상과 로봇수술 동영상에 딥러닝 기술을 적용하고 그

유효성을 확인하였으며, 향후에 제안하는 방법이 임상에서 사용되고 있는 진단 및 평가 방법의 대안이 될 것으로 기대한다.

주요어: 딥러닝, 합성곱 신경망, 대장내시경 검사, 수술도구 추적, 로봇수술 술기 평가

학 번: 2015-30264

Acknowledgement

박사학위 과정 동안에 감사한 모든 분들께 이 글을 바칩니다. 가장 먼저 지도 교수님이신 김희찬 교수님께 진심으로 감사의 말씀을 드립니다. 여러모로 부족했던 저를 지도해주시고 성장할 수 있도록 많은 기회들을 주셨습니다. 다만 연구 뿐만 아니라, 사람으로서 정도(正道)를 갈 수 있도록 이끌어주신 가르침을 항상 가슴에 새기겠습니다.

수업과 연구 외에도 인간적인 따뜻함으로 대해주신 의공학과의 윤형진 교수님과 이정찬 교수님께도 감사를 드립니다. 연구실 선배님이자 연구와 실무를 가르쳐주신 공현중 교수님, 그리고 관심과 배려로 많은 도움을 주신 구윤서 교수님과 이사람 교수님께도 감사를 드립니다.

내분비외과의 이규언 교수님, 채영준 교수님, 유형원 교수님께도 감사를 드립니다. 저의 부족함에도 오랫동안 지지해주셨기에 학문적으로 더 깊어질 수 있었던 것 같습니다. 좋은 연구 성과를 낼 수 있도록 도움을 주신 소화기내과의 김주성, 이창현, 진은효 교수님께도 감사를 드립니다. 또한 영상의학과의 박창민 교수님, 김형진 교수님, 피부과의 정진호 교수님, 이동훈 교수님, 나정임 교수님, 윤현선 교수님, 조수익 교수님, 재활의학과의 오병모 교수님, 서한길 교수님, 이우형 교수님, 비뇨기과의 오승준 교수님, 소아외과의 김현영 교수님, 윤중기 교수님께도 감사를 드립니다.

여러 임상과 선생님들과의 공동연구는 소중한 경험이었고 협업과 소통의 중요성, 그리고 의학적인 관점들을 배울 수 있었습니다.

동고동락하며 힘이 되어준 MELAB 의 석규, 승만, 희안, 지은, 준녕, 치헌, 희진이형, 장재, 순빈이와 졸업한 선배님들께도 감사의 인사를 전합니다. 또한 MBDL 의 동아, 경진, 윤하형, 찬훈, 우상, 태우, 준희에게도 고맙다는 말을 하고 싶습니다.

학위 과정은 치열하게 고민하는 시간들의 연속이었습니다. 하나의 연구 주제가 한 편의 논문으로 완성되기까지 많은 시간과 고뇌가 요구됩니다. 그러나 이러한 과정 끝에 진정 남는 것은 한 명의 훈련된 연구자일 것입니다. 졸업식이라는 단어는 '시작 (Commencement)' 이라는 의미가 있습니다. 그동안에 얻은 지식과 경험, 더 커진 역량으로 다시 새로운 연구를 시작할 생각에 가슴이 뛥니다!

마지막으로 바쁜 학위 과정을 이해해주고 배려해준 아내에게 고맙고 사랑한다는 말을 전합니다. 사랑하는 가족들, 그리고 조건없이 목양해준 교회와 항상 인도하시는 하나님께 감사를 드립니다.