COMPUTER VISION-BASED NEURO-ENDOSCOPIC SURGICAL VIDEO ANALYSIS AND EVALUATION SYSTEMS

BRITTY BABY

AMAR NATH AND SHASHI KHOSLA SCHOOL OF INFORMATION TECHNOLOGY
INDIAN INSTITUTE OF TECHNOLOGY DELHI
DECEMBER 2022

COMPUTER VISION-BASED NEURO-ENDOSCOPIC SURGICAL VIDEO ANALYSIS AND EVALUATION SYSTEMS

by

Britty Baby

AMAR NATH AND SHASHI KHOSLA SCHOOL OF INFORMATION TECHNOLOGY

Submitted

in fulfillment of the requirements of the degree of Doctor of Philosophy

to the

INDIAN INSTITUTE OF TECHNOLOGY DELHI

DECEMBER 2022

THESIS CERTIFICATE

This is to certify that the thesis titled **Computer Vision based Neuro-endoscopic Surgical Video Analysis and Evaluation Systems**, submitted by **Britty Baby (2014ANZ8198)**, to the Indian Institute of Technology, Delhi, for the award of the degree of **Doctor of Philosophy**, is a bona fide record of the research work done by her under our supervision. The contents of this thesis, in full or in parts, have not been submitted to any other Institute or University for the award of any degree or diploma.

Certain works included in this thesis involved collaboration by other students, which have been explicitly specified/acknowledged in the corresponding chapters and the part done by those collaborators appeared in their respective reports/thesis.

Prof. Subhashis Banerjee
Professor
Department of Computer
Science and Engineering
Indian Institute of Technology Delhi

Prof. Ashish Suri
Professor
Department of Neurosurgery
All India Institute
of Medical Sciences
New Delhi

Prof. Chetan Arora
Associate Professor
Department of Computer
Science and Engineering
Indian Institute of Technology Delhi

Place: New Delhi
Date: 20 December 2022

ACKNOWLEDGEMENTS

"Gratitute is not something you have to express. If you are filled with gratitude for all things that contribute your life, it will melt your very being."

—**Sadhguru**

I am extremely grateful to all that makes my life happen. I want to express my gratitude to all of my supervisors who encouraged me to contribute everything I could to translational research of the country's best universities, IIT Delhi and AIIMS.

I want to offer my gratitude to Prof. Subhashis Banerjee, my supervisor, for his advice and assistance during my study. He always offered me the flexibility to do my own study and provided crucial assistance and guidance anytime I was stuck in the plethora of information.

I offer my gratitude to Prof. Chetan Arora, for his mentorship and assistance in my development as a computer vision researcher. He has always been available to provide constructive criticism and supported with involving other research students in order to make this work an interactive and educational experience for me. He motivated me to push the boundaries of my comfort zone, to go further into the research, and tap into a capacity I never imagined I had.

I am forever grateful to my supervisor Prof. Ashish Suri for being a mentor and the foundation force to make this translational research happen. It is his vision of neurosurgery skills evaluation that got manifested through me. Without his encouragement and direction, I would not have been able to go on this path of research and professional life.

Additionally, I would like to express my gratitude to Prof. Prem K Kalra for his assistance and insightful recommendations during my study. I would like to express my gratitude to Prof. Subodh Kumar and Prof. Rahul Narain for their assistance and prompt direction, as well as several recommendations during the research meetings. I am grateful to Prof. Sumantra Dutta Roy of the Electrical Engineering Department for his encouragement and suggestions during my study.

I would like to convey my appreciation to my collaborators Daksh Thapar, a research scholar at IIT Mandi, Mustafa Chasmai, and Tamajit Banerjee, undergraduate students at IIT Delhi, who assisted me with the experiments. Additionally, I would like to thank Kunal Dargan, Rohan Raju Dhanakshirur, and Argha Chakraborty, research students at the School of Information Technology, as well as Dr. Ramandeep Singh, Dr. Vinkle Srivastav

and Rajdeep Singh of the Neurosurgery Education and Training School (NETS), All India Institute of Medical Sciences (AIIMS), for their assistance with my research. I would want to convey my thanks to my colleagues and other staff members at the Medical Lab, Vision Lab, and NETS for their contributions to my study, whether directly or indirectly.

I appreciate my parents and family members' encouragement, support, and collaboration. I would want to pay tribute to my late brother, who always supported and encouraged me to pursue my career.

I offer immense gratitude to Sadhguru Jagadish Vasudev; I am not sure whether I would be where I am now without his grace and transformational yogic tools, which have helped me sail over the most challenging circumstances in my life to date.

Britty Baby

ABSTRACT

Minimally invasive neurosurgery involves a surgical intervention to treat the diseases of the brain and spinal cord by smaller incisions or reaching through natural orifices with the help of an endoscope. Iatrogenic errors that occur while performing the surgery account for the eighth leading cause of death in America. This has drawn attention to training surgeons in fundamental technical skills. The neuro-endoscopic procedures are complex because of physical, visual, motor, spatial, and haptic constraints and minimal margin of error. The existing skills training method through the apprenticeship model is time-bound (3yrs/6yrs), mentor-dependent, non-uniform, anecdotal, and slow skills acquisition. This thesis is motivated by the fact that there are no validated skills evaluation methods in Neuroendoscopy. Our contributions are based on the collaborative efforts of surgeons and technologists to provide computer vision-based standardized and automatic skills evaluation modules to improve neuro-endoscopic skills training.

We started with a systematic review of the existing skills training modules available in Neuroendoscopy including the virtual and physical simulation models and their evaluation methods. We found that physical simulators provide real surgical hands-on skills exposure and are cost-effective simulation models that can facilitate video analysis of the simulation. We used a novel open-source physical partial task simulator called a Neuro-Endo-Trainer (`NET`) developed by our group for our evaluation studies.

In this thesis, we incorporate an auxiliary static camera inside the `NET` training box to record and analyze the surgical activity to facilitate automatic activity detection. The video analysis is performed by automatically segmenting the input video stream into sub-tasks using a Mixture of Gaussian-based background subtraction and the Tracking Learning Detection algorithm. The motion statistics in each sub-task are collated into a synopsis to provide feedback to the trainee.

Surgical instruments are made of stainless steel and traditional computer vision-based tracking models fail to track the tool robustly. We explore deep-learning-based tool segmentation methods and more sophisticated machine learning-based skills evaluation methods. For this, we propose new datasets; the data of `NET` simulation environment as Neuro-endoscopic Technological Skills Training Dataset (`NETS`) and clinical neuro-endoscopic endonasal endoscopic transsphenoidal surgery (`EETS`). We also analyze the various datasets available in the literature for surgical instrument segmentation and skills training like En-

doscopic Vision Challenge EndoVis2017 (EV17), Endovis2018 (EV18), and JHU-ISI Gesture and Skill Assessment Working Set (JIGSAWS). We augment instance segmentation to EV18 and JIGSAWS datasets. We then explore various deep-learning methods for accurate instance segmentation of surgical instruments.

We identified that current state-of-the-art (SOTA) techniques fine-tune contemporary models trained on natural images for surgical instance segmentation, but manage to give a low accuracy only (Challenge IOU = 0.55). Our investigation reveals that, though the accuracy of the bounding box and mask is high, the classification head performs poorly. We perform strategic modifications in the modern two-stage instance segmentation model to introduce a third stage specializing in classification to correct the prediction masks labels. We introduce multi-scale mask attention in the third stage to emphasize the instrument regions in the image. Surgical instruments are visually very similar, so typical cross entropy-based classifier training is insufficient. We train our third stage using metric learning and arcloss to correctly classify the predicted mask instance. Compared with almost 18 methods published in top vision conferences, our exhaustive experiments show that we improve upon all the methods and achieve at least 17 points (30%) improvement over the SOTA on the benchmark EV17 challenge. We conduct exhaustive experiments on the other datasets EV18 and the proposed EETS dataset. We also propose a novel Cumulative IOU metric for instance segmentation of instruments.

We then explore machine-learning-based evaluation methods to improve our synopsis-based skills evaluation framework of NET. We extend the feedback to the trainee based on machine-learning based adaptation of the expert evaluation to provide a robust, automatic, and repeatable technological solution for surgical skills evaluation. A multi-path neural network framework for automated skills assessment using various surgical skill aspects is a promising and adaptable solution for different surgical scenarios. We study the multi-path framework for skills assessment for JIGSAWS and proposed NETS dataset. Our experiments show that the method does not generalize well to our dataset. Hence, we propose a dynamic variant of rank-loss for contrastive learning of video representation along with the Mean-Squared Error (MSE) loss. This enables the network to learn the discriminative features that predict the relative rank of the skills for pairs and thereby better correlate with the expert surgeon's ranking. We provide cutting edge results on surgical skills evaluation for JIGSAWS and proposed NETS dataset with improvement of 5% and 27 % respectively over SOTA method. It is expected that the outcomes of this research will pave the path for developing more sophisticated video analysis algorithms, hybrid and augmented reality simulators, and automated skills evaluation methods in the field of neurosurgery.

KEYWORDS: Neuroendoscopy, Simulators, Video Analysis, Activity Detection, Surgical Instrument, Instance Segmentation, Automatic skills Evaluation

सार

मिनिमली इनवेसिव न्यूरोसर्जरी एक सर्जिकल हस्तक्षेप है जिसमें मस्तिष्क और रीढ़ की हड्डी के रोगों के इलाज के लिए छोटे चीरों से या प्राकृतिक छिद्रों से एंडोस्कोप की मदद से पहुंचने का प्रयास करते हैं। सर्जरी के दौरान होने वाली आईट्रोजेनिक त्रुटियां, अमेरिका में मृत्यु के आठवें प्रमुख कारण हैं। इसने मौलिक तकनीकी कौशल में सर्जनों को प्रशिक्षित करने पर ध्यान आकर्षित किया है। न्यूरो-एंडोस्कोपिक प्रक्रियाएं भौतिक, दृश्य, मोटर, स्थानिक और हैप्टिक बाधाओं और त्रुटि के न्यूनतम मार्जिन के कारण जटिल हैं। अप्रेंटिसशिप मॉडल के माध्यम से मौजूदा कौशल प्रशिक्षण पद्धति समयबद्ध (3 वर्ष / 6 वर्ष), संरक्षक-निर्भर, गैर-समान, वास्तविकऔर धीमी गति से कौशल अधिग्रहण है। यह थीसिस इस तथ्य से प्रेरित है कि न्यूरोएन्डोस्कोपी में कोई मान्य कौशल मूल्यांकन विधियां नहीं हैं। हमारा योगदान न्यूरो-एंडोस्कोपिक कौशल प्रशिक्षण में सुधार के लिए कंप्यूटर विज़न-आधारित मानकीकृत और स्वचालित कौशल मूल्यांकन मॉड्यूल प्रदान करने के लिए सर्जनों और टेक्नोलॉजिस्ट्स के सहयोगात्मक प्रयासों पर आधारित है।

हमने वर्चुअल और फिजिकल सिमुलेशन मॉडल और उनके मूल्यांकन विधियों सहित न्यूरोएंडोस्कोपी में उपलब्ध मौजूदा कौशल प्रशिक्षण मॉड्यूल की एक व्यवस्थित समीक्षा के साथ शुरुआत की। हमने पाया कि फिजिकल सिमुलेटर वास्तविक सर्जिकल हैंड्स-ऑन कौशल प्रदर्शन प्रदान करते हैं और लागत प्रभावी सिमुलेशन मॉडल हैं जो सिमुलेशन के वीडियो विश्लेषण की सुविधा प्रदान कर सकते हैं। हमने अपने मूल्यांकन अध्ययनों के लिए हमारे समूह द्वारा विकसित न्यूरो-एंडो-ट्रेनर (NET) नामक एक उपन्यास ओपन-सोर्स फिजिकल पार्शियल टास्क सिम्युलेटर का उपयोग किया।

इस थीसिस में, हमने NET प्रशिक्षण बॉक्स के अंदर एक सहायक स्थैतिक कैमरा शामिल किये हैं जिससे हम स्वचालित गतिविधि का पता लगाने के लिए सर्जिकल गतिविधि को रिकॉर्ड और विश्लेषण करते है।गॉसियन-आधारित बैकग्राउंड सेगमेंटेशन और ट्रैकिंग लर्निंग डिटेक्शन एल्गोरिदम के मिश्रण का उपयोग करके इनपुट वीडियो स्ट्रीम को स्वचालित रूप से उप-कार्यों में विभाजित करके वीडियो विश्लेषण किया जाता है। प्रशिक्षु को प्रतिक्रिया प्रदान करने के लिए प्रत्येक उप-कार्य में गति के आँकड़ों को एक सिनोप्सिस में जोड़ा जाता है।

सर्जिकल उपकरण स्टेनलेस स्टील से बने होते हैं और पारंपरिक कंप्यूटर दृष्टि-आधारित ट्रैकिंग मॉडल उपकरण को मजबूती से ट्रैक करने में विफल होते हैं। हम डीप-लर्निंग-आधारित टूल सेगमेंटेशन विधियों और अधिक परिष्कृत मशीन लर्निंग-आधारित कौशल मूल्यांकन विधियों का पता लगाते हैं। इसके लिए, हमने नए न्यूरो-एंडोस्कोपिक टेक्नोलॉजिकल स्किल्स ट्रेनिंग डेटासेट (NETS) के रूप में NET सिमुलेशन वातावरण का डेटा और क्लिनिकल न्यूरो-एंडोस्कोपि के एंडोनासल एंडोस्कोपिक ट्रांसफेनोइडल सर्जरी (EETS) डेटा प्रस्तावित किये हैं। हम सर्जिकल इंस्ट्रूमेंट सेगमेंटेशन और कौशल प्रशिक्षण के लिए साहित्य में उपलब्ध विभिन्न डेटासेट जैसे एंडोस्कोपिक विजन चैलेंज एंडोविस2017 (EV17), एंडोविस2018 (EV18), और JHU-ISI जेस्चर और स्किल असेसमेंट वर्किंग सेट (JIGSAWS) का भी विश्लेषण करते हैं। हम EV18 और JIGSAWS डेटासेट में इंस्टैंस सेगमेंटेशन डेटा जोड़के उनके डाटासेट बढ़ाये हैं। फिर हम न्यूरोसर्जरी उपकरणों के सटीक इंस्टैंस सेगमेंटेशन के लिए विभिन्न गहन-शिक्षण विधियों का पता लगाते हैं।

हमने पहचाना की कि वर्तमान अत्याधुनिक (SOTA) तकनीक सर्जिकल इंस्टैंस सेगमेंटेशन के लिए प्राकृतिक छवियों पर प्रशिक्षित समकालीन मॉडलों को फ़ाइन ट्यून करती है, लेकिन केवल कम सटीकता देने का प्रबंधन करती है (चैलेंज IOU = 0.55)। हमारी जांच से पता चलता है कि हालांकि बाउंडिंग बॉक्स और मास्क की सटीकता अधिक है, लेकिन क्लासिफिकेशन हेड खराब प्रदर्शन करता है। हम अनुमानित मास्क लेबल को सही करने के लिए क्लासिफिकेशन के तीसरे चरण को पेश करते हुए आधुनिक दो-चरण इंस्टैंस सेगमेंटेशन मॉडल में एक रणनीतिक संशोधन करते हैं। हम छवि में साधन क्षेत्रों पर जोर देने के लिए तीसरे चरण में बहु-स्तरीय मास्क अटेंशन पेश करते हैं। सर्जिकल उपकरण दृष्टिगत रूप से बहुत समान हैं, इसलिए विशिष्ट क्रॉस एन्ट्रापी-आधारित क्लासिफायरियर प्रशिक्षण अपर्याप्त है। हम अनुमानित मास्क उदाहरण को सही ढंग से वर्गीकृत करने के लिए मीट्रिक लर्निंग और आर्कलॉस का उपयोग करके अपने तीसरे चरण को प्रशिक्षित करते हैं। शीर्ष दृष्टि सम्मेलनों में प्रकाशित लगभग 18 विधियों की तुलना में, हमारे संपूर्ण प्रयोग बताते हैं कि हम सभी तरीकों में सुधार करते हैं और बेंचमार्क EV17 चुनौती पर SOTA पर कम से कम 17 अंक (30%) सुधार प्राप्त करते हैं। हम अन्य डेटासेट EV18 और प्रस्तावित EETS डेटासेट पर संपूर्ण प्रयोग करते हैं।हम उपकरणों के इंस्टैंस सेगमेंटेशन के लिए एक नई क्युमुलेटिव IOU मीट्रिक भी प्रस्तावित करते हैं।

फिर हम NET के अपने पहले के सिनोप्सिस-आधारित कौशल मूल्यांकन ढांचे में सुधार के लिए मशीन-लर्निंग-आधारित मूल्यांकन विधियों का पता लगाते हैं। हम एक्सपर्ट मूल्यांकन के मशीन-लर्निंग आधारित

अनुकूलन के आधार पर एक मजबूत, स्वचालित और दोहराने योग्य तकनीकी समाधान प्रदान करते हुए प्रशिक्षु के कौशल मूल्यांकन को प्रतिक्रिया देते हैं। एक मल्टी-पाथ न्यूरल नेटवर्क फ्रेमवर्क हैं जिसमें विभिन्न सर्जिकल कौशल पहलुओं का उपयोग किये हैं, वे विभिन्न सर्जिकल परिदृश्यों के स्वचालित कौशल मूल्यांकन के लिए एक आशाजनक और अनुकूलनीय समाधान है। हम JIGSAWS और प्रस्तावित NETS डेटासेट के कौशल मूल्यांकन के लिए मल्टी-पाथ फ्रेमवर्क का अध्ययन करते हैं। हमारे प्रयोग बताते हैं कि यह विधि हमारे डेटासेट के लिए अच्छी तरह से सामान्यीकृत नहीं होती है। इसलिए, हम मीन-स्क्वेर्ड एरर (MSE) लॉस के साथ-साथ वीडियो प्रतिनिधित्व के कॉन्ट्रास्टिव लर्निंग के लिए रैंक-लॉस का डायनेमिक वेरिएंट प्रस्तावित करते हैं। यह नेटवर्क को उन भेदभावपूर्ण विशेषताओं को सीखने में सक्षम बनाता है जो वीडियो के एक जोड़े के लिए कौशल के सापेक्ष रैंक की भविष्यवाणी करते हैं और इस तरह एक्सपर्ट सर्जन की रैंकिंग के साथ बेहतर संबंध स्थापित करते हैं। हम SOTA पद्धति पर क्रमश: 5% और 27 % के सुधार के साथ JIGSAWS और प्रस्तावित NET डेटासेट के लिए सर्जिकल कौशल मूल्यांकन पर अत्याधुनिक परिणाम प्रदान करते हैं। यह उम्मीद की जाती है कि इस शोध के नतीजे न्यूरोसर्जरी के क्षेत्र में अधिक परिष्कृत वीडियो विश्लेषण एल्गोरिदम, हाइब्रिड और ऑगमेंटेड रियलिटी सिमुलेटर और स्वचालित कौशल मूल्यांकन विधियों के विकास के लिए मार्ग प्रशस्त करेंगे।

कीवर्ड्स: न्यूरोइंडोस्कोपी, सिमुलेटर, वीडियो एनालिसिस, एक्टिविटी डिटेक्शन, सर्जिकल इंस्ट्रूमेंट, इंस्टैंस सेगमेंटेशन , आटोमेटिक स्किल्स इवैल्यूएशन

Contents

List of Figures

List of Tables

ABBREVIATIONS

2D	2-Dimensional
3D	3-Dimensional
OR	Operating Room
VR	Virtual Reality
ETV	Endoscopic Third Ventriculostomy
EETS	Endoscopic Endonasal Transsphenoidal Surgery
NET	Neuro-Endo-Trainer
OSATS	Objective structured assessment of technical skill
SIMONT	Sinus Model Oto-Rhino Neuro Trainer
A.S.P.E.N.	Anatomical Simulator for Pediatric Neurosurgery
NEVAT	Neuroendoscopic Ventriculostomy Assessment Tool
CL	Checklist
GRS	Global Rating Scale
LMIC	Low- and Middle-Income Countries
ESS	Endoscopic Sinus Surgery Simulator
EEA	Endoscopic Endonasal Approach
CT	Computed Tomography
USB	Universal Serial Bus
ICSAD	Imperial College Surgical Assessment Device
IMU	Inertial Measurement Unit
NETS-SAS	Neurosurgery Education and Training School- Skills Assessment Scale
TLD	Tracking Learning Detection algorithm
HOG	Histogram of Oriented Gradients
LSVM	Latent Support Vector Machine
SVM	Support Vector Machine
CCA	Canonical correlation analysis
ENTS	Endo-Nasal Transsphenoidal Skull-base surgeries
VE	Virtual Endoscopy
3D	3-Dimensional
VIVENDI	Virtual Ventricle Endoscopy
MRI	Magnetic Resonance Imaging
MRA	Magnetic Resonance Angiography
KisMo	Kismet modeler

KFF	Kismet Force Feedback
FEM	Finite Element Method
STL	Stereolithography
VTK	Visualization Toolkit
VVRS	Visualization Virtual Reality Simulation
SR-VE	Surface Rendered-Virtual Endoscopy
VR-VE	Volume Rendered-Virtual Endoscopy
PPU	Physics Processing Unit
SPH	Smoothed Particle Hydrodynamics
FESS	Fundamental Endoscopic Sinus Surgery Training Simulator
DOF	Degrees of Freedom
NES	Nasal Endoscopy Simulator
CT	Computed Tomography
ESS or ES3	Endoscopic Sinus Surgery Simulator
ENT	Ear Nose Throat
MISTVR	Minimally Invasive Surgical Trainer Virtual Reality
PicSOr	Pictorial Surface Orientation
VSE	Virtual Surgical Environment
ITK	Image Processing ToolKit
OHNS	Otorhinolaryngology-Head & Neck Surgery
RMOs	Resident Medical Officers
MSESS	McGill simulator for endoscopic sinus surgery
ECE	Educational Computer-based-simulation Environment
AR	Augmented Reality
NRC	National Research Council of Canada
GAINS	Global Assessment of Intraoperative Neurosurgical Skills
NEVAT	Neuro-Endoscopic Ventriculostomy Assessment Tool
JIGSAWS	JHU-ISI Gesture and Skill Assessment Working Set
EndoVis17	Endoscopic Vision Challenge 2017
EndoVis18	Endoscopic Vision Challenge 2018
EndoVis19	Endoscopic Vision Challenge 2019
NETS-SAS	Neurosurgery Education and Training School-Skills Assessment Scale
RobustMIS	Robust Medical Instance Segmentation 2019
CADIS	Cataract Dataset for Image Segmentation
TLD	Tracking-Learning-Detection
NETS	Neuro-endoscopic Technical Skills Training Dataset
ISINet	Instance-based Surgical Instrument Segmentation Network
NETA	Neuro-EndoTrainer Activity Dataset
NETE	Neuro-EndoTrainer Event Dataset
NETIS	Neuro-EndoTrainer Instrument Segmentation

COCO Common-Objects in Context