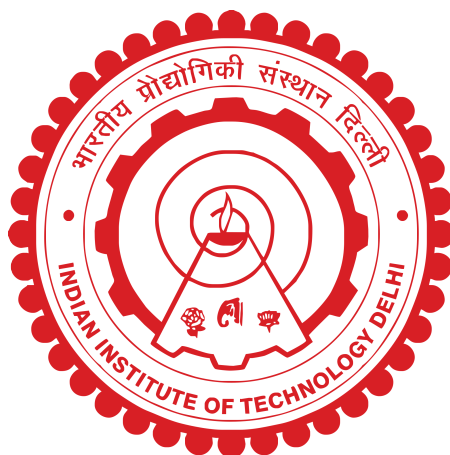


COMPUTER VISION-BASED
NEURO-ENDOSCOPIC SURGICAL VIDEO
ANALYSIS AND EVALUATION SYSTEMS

BRITTY BABY



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

© Indian Institute of Technology Delhi (IITD), New Delhi, 2022

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

by

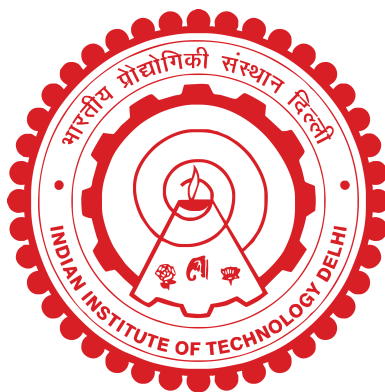
Britty Baby

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

“Gratitude 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-

oscopic 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, वास्तविक और धीमी गति से कौशल अधिग्रहण है। यह थीसिस इस तथ्य से प्रेरित है कि न्यूरोएंडोस्कोपी में कोई मान्य कौशल मूल्यांकन विधियां नहीं हैं। हमारा योगदान न्यूरो-एंडोस्कोपिक कौशल प्रशिक्षण में सुधार के लिए कंप्यूटर विज्ञान-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 सिमुलेटर 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 हैं। गॉसियन-आधारित बैकग्राउंड सेगमेंटेशन और ट्रेकिंग 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-एंडोस्कोपिक 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 क्लासिफायरिपर 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 व्युत्प्रेतव 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 है 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 क्रमशः 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 न्यूरसर्जरी 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

ACKNOWLEDGEMENTS	1
ABSTRACT	3
LIST OF FIGURES	15
LIST OF TABLES	17
ABBREVIATIONS	18
1 Introduction	1
1.1 Endoscope and Optics	1
1.2 Neuroendoscopy	2
1.2.1 Neuroendoscopy challenges	3
1.2.2 Neuroendoscopy Skills Training	5
1.2.3 Neuroendoscopy Video Analysis	7
1.3 Simulators & Skills Evaluation Methods	7
1.4 Summary of contributions	9
1.4.1 Surgical Activity Detection	9
1.4.2 Surgical Activity Data Collection and Annotations	11
1.4.3 Surgical Tool Segmentation	12
1.4.4 Surgical Skills Evaluation	12
1.5 Organization	13
2 Literature Survey	14

- 2.1 Virtual Reality Simulators 15
 - 2.1.1 Introduction 15
 - 2.1.2 Results 16
 - 2.1.3 Discussion 25
 - 2.1.4 Conclusion 31
- 2.2 Physical Simulators 32
 - 2.2.1 Introduction 32
 - 2.2.2 Results 33
 - 2.2.3 Discussion 39
 - 2.2.4 Conclusion 42
- 2.3 Simulators and Video Analysis: A perspective towards the relevance 42
- 3 Surgical Activity Detection 44**
 - 3.1 Background 45
 - 3.1.1 Novel Physical Simulator: NET 45
 - 3.1.2 Validation and learning from the simulator 47
 - 3.2 Neuro-Endoscopic Activity Tracker 48
 - 3.2.1 Introduction 48
 - 3.2.2 Related Work 49
 - 3.2.3 Methods 51
 - 3.2.4 Results 58
 - 3.2.5 Conclusion 61
- 4 Surgical Activity Data Collection and Annotations 62**
 - 4.1 Introduction 62
 - 4.2 Developed Datasets 63
 - 4.2.1 Physical Simulator Skills Assessment: NET Dataset 63
 - 4.2.2 Instance Segmentation Datasets 63

CONTENTS	10
4.3 Related Robotic Surgery Datasets	65
4.3.1 Augmentation of Public Datasets	67
4.4 Conclusion	68
5 Surgical Tool Segmentation	69
5.1 Introduction	69
5.2 Background and Motivation	72
5.3 Proposed Three Stage Model (S3NET)	73
5.3.1 Post-processing of second stage output	74
5.3.2 Handling misclassification	74
5.3.3 Multi Scale Mask Attended (MSMA) Classifier	75
5.3.4 Training	77
5.4 Dataset and Evaluation	77
5.5 Experiments and Results	81
5.5.1 Ablation studies	81
5.6 Metric For Instance Segmentation	89
5.7 Conclusion	90
6 Surgical Skills Evaluation	92
6.1 Introduction	92
6.2 Proposed Methodology	94
6.2.1 Unified Skill Evaluation	94
6.2.2 K-Rank loss	95
6.3 Experiments and Results	97
6.3.1 Experimental Setup	97
6.3.2 Training Hyperparameters	97
6.3.3 Results on JIGSAWS Dataset	98
6.3.4 Results on NETS Dataset	98

6.4	Conclusion	99
7	Conclusion and Future Directions	100
7.1	Summary of contributions	100
7.2	Future Perspectives	102
7.2.1	Endoscopic video skills evaluation	102
7.2.2	Video-Instance segmentation of Surgical Instruments	103
7.2.3	Procedure Simulators	103
7.2.4	Evaluation of surgeon's parameters	103
7.2.5	Skills evaluation using spatio-temporal networks	103
7.2.6	3D tracking and Neuro-navigation simulator	104
	Bibliography	105
	LIST OF PAPERS BASED ON THESIS	132
	BIO-DATA	134

List of Figures

1.1	(A) High definition Neuro-endoscope system Neuro-endoscope and instruments: B: Ventricular Lotta endoscope and instrument set C: Ventricular DECQ endoscope and instrument set, D: Endo-nasal Rigid endoscopes (0^0 , 30^0 , 45^0), E: Endo-nasal instrument set	2
1.2	[ETV]: (A) Planning on magnetic resonance imaging. (B) Burr-hole incision and placement of endoscope. (C) Endoscopic view of the third ventriculostomy	3
1.3	[EETS]:(A) Planning on magnetic resonance imaging. (B) Patient positioning for endonasal approach. (C) Endoscopic view of endonasal transsphenoidal surgery	4
1.4	Illustration of (A) microscopic surgery with a stereo vision and 3D view of the surgical site and (B) endoscopic surgery with 2D visual feedback and fulcrum effect	4
1.5	Illustration of endoscopic images of endo-nasal transphenoidal skull-base surgeries (A, B, C, D, E, F). (It is difficult to perceive the shape of the object in the images because of small field of view, reflections, partial boundaries, distortion, lack of texture, and shading caused by near-field lighting.) . . .	5
1.6	Graphs to show the neuro-endoscopic learning curve pre and post training on virtual task. (Upper) For novices (Lower) For experienced surgeons. White bars for pre-training and Grey bars post-training [HFU ⁺ 13a]	6
1.7	Methods available in literature for skills evaluation (A) IMU-based sensors [LZS ⁺ 09] (B) Tr-Endo tracking system [CBGD06] (C) Error Detection modules [ECdLM15] (D) Video-based tool tracking [GAH08]	6
1.8	Summary of thesis contributions	10
1.9	Various stages of neuro-endoscopy surgery covered in EETS dataset	12
2.1	Classification of neuroendoscopic virtual simulators for endoscopic third ventriculostomy and endoscopic endo-nasal surgeries.	15
2.2	Neuro-Touch virtual reality simulator: (A) Complete training setup, (B) ETV surgery view on simulator and (C) EETS surgery view on simulator	23

2.3	Immersive Touch virtual reality simulator showing the user console and display	25
2.4	Classification of physical simulators for neuroendoscopic ventricular and endonasal surgeries.	32
2.5	Generalized training setup of synthetic simulators	34
2.6	Generalized training setup of box trainers	38
3.1	(A) Trainee using the physical simulator designed by our group in the Neurosurgery Skills Training Facility, AIIMS, (B) NET box (C) Endoscopic display of the activity.	44
3.2	Training pattern on activity plate of the endo-trainer: A: To train with horizontal movements, the rings on the right side are moved to the left as shown by the green arrows. B: To help in training for diagonal movements, the rings are placed on the diagonally opposite pegs as indicated by the white arrows.	46
3.3	Neuro-Endo-Trainer with auxiliary camera attached to the top.	48
3.4	State-machine of the Neuro-endoscopy activity tracker	54
3.5	Flow chart of endo-tracker algorithm	56
3.6	(A) Peg segmentation output;(B) Ring segmentation output	57
3.7	Activity Detection output of state machine at every frame	58
3.8	Screenshot of acquisition software for endoscope and auxiliary camera.	59
3.9	Sample synopsis of the activity	60
4.1	(A) A frame from NET dataset (B) Instance segmentation output of the frame	64
4.2	Instrument samples from EETS dataset. First row, left to right: Suction, Irrigation, Dissector, Scissors, and Knife. Second row, left to right: Navigation, Biopsy, Curette, Drill, and Tumor_biopsy.	65
4.3	Snapshots from EV17 dataset dataset [Endb]	66
4.4	Snapshots from the tasks in JIGSAWS (suturing, knot tying, and needle passing) [GVR ⁺ 14]	67

5.1	Instrument segmentation produced by various competitive methods on a sample from the EV17 dataset [Endb]. Each instrument class is shown in a different color. Note that ISINet gets the segmentation right but classifies incorrectly. We identify instrument misclassification as the primary reason for the low performance of the SOTA techniques. To indicate the severity of the problem, we just change the predicted class label of an object predicted by MaskRCNN with the ground truth label and AP50 of the model improves from 0.65 to 0.90 on the dataset.	70
5.2	The figure shows the proposed 3-Stage Network (S3NET). Whereas the first two stages are similar to state of the art, we introduce a third stage, MSMA, specializing in classification. We make several innovations in the design of MSMA as described in the main text.	73
5.3	Qualitative Analysis for the comparison of instance segmentation: 4 symbols are used to show the results; represents the instance labeled correctly, shows the misclassified instance, and ‘O’ represents missed instance, A letter ‘A’ indicates an ambiguous instance, where it is ambiguous to select the class of the instrument either due to over-segmentation or due to multiple copies of instrument classes at the same region. We show better classification in the cases of sparse class, overlapping instruments, and do not miss instrument instances. Our failure cases include cases where the instance shows the only shaft of the instrument and a significant change in instrument orientation.	80
5.4	Percentage Distribution of classes in the EV17 train dataset	82
5.5	Qualitative Analysis of Classification of Bipolar Forceps. Failure includes when the instrument articulation changes	83
5.6	Qualitative Analysis of Classification of Prograsp Forceps. Failure includes when the input mask contains a combined instances	83
5.7	Qualitative Analysis of Classification of Large Needle Driver	84
5.8	Qualitative Analysis of Classification of Vessel Sealer. Failure case includes when the tip is zoomed in	84
5.9	Qualitative Analysis of Classification of Grasping Retractor. Failure case includes when only the shaft is visible in the frame.	85
5.10	Qualitative Analysis of Classification of Monopolar Curved Scissors. Failure case include when the mask is corrupted	85
5.11	Qualitative Analysis of Classification of Ultrasound Probe. Failure case include when the mask is corrupted with no instances identified.	86
6.1	Block diagram of a general automated surgical skills assessment system. The top, and bottom rows show samples from JIGSAWS, and NETS dataset respectively.	93

6.2	An overview of the model architecture for skill evaluation. Different features are passed through different paths in the network followed by a path dependency module. The scores are combined using the dependency module as given by Eq. (6.1)	94
6.3	An example highlighting differences between MSE and SROCC. In the given figure, the MSE will be very good because of small absolute differences. On the other hand, SROCC will be poor because relative ranking between samples 1 and 2 is inconsistent.	96
7.1	3D model of the skull	104
7.2	Setup for visualization of CT points with respect to tracking of tool	105

List of Tables

2.1	Comparison of Virtual Reality Simulators for ETV (Visualization)	26
2.2	Comparison of Virtual Reality Simulators for ETV (Haptics Interaction) .	27
2.3	Comparison of Virtual Reality Simulators for EETS surgery (Visualization)	27
2.4	Comparison of Virtual Reality Simulators for EETS surgery (Haptics Interaction)	28
2.5	Detailed classification of neuro-endoscopic virtual skills training simulators, their assessment methods and validation results	28
2.6	Detailed classification of neuro-endoscopic physical skills training simulators, their assessment methods and validation results	41
3.1	Skills Assessment Scale	45
3.2	Comparison of skills assessment scale score and task completion time using a 0° scope and straight plate amongst Groups E, N, and R*(*Group R data taken from the first iteration only)	46
3.3	Objective Measure equivalent to NETS-SAS from video stream	49
3.4	NETS-SAS-Video parameters and their respective states	57
3.5	Evaluation results of a set of experts and trainees	59
5.1	Results for 4-fold validation with train-test split as described in the paper. We report results in terms of challenge IOU metric as suggested in EV17. SOTA <i>NLI methods</i> , are general purpose instance segmentation techniques which have appeared in last few years and have reported their results on natural images. For the purpose of comparison we have fine-tuned them on EV17. EVS have been specifically proposed for medical instrument segmentation. .	75

5.2	In Tab. 5.1 we have compared with SOTA methods after fine-tuning them on EV17. However, we propose a specific post-processing specially for the instrument segmentation scenario which does a non-maximal suppression across the classes. Such post-processing can be applied to other methods as well, and hence for a fair comparison we have done so. The table shows the results after post-processing. Note that other EVS methods already do a similar post-processing along with host of others. Hence their results do not change. Note that, we do not claim novelty/contribution over the suggested post-processing, as it is already implemented in EVS methods without formally describing/acknowledging it.	76
5.3	Performance of SOTA instance segmentation methods on EV17 and EV18 instrument segmentation datasets. (R50 represents ResNet-50-FPN, Trfmr represents Transformer, BF-Bipolar Forceps, PF-Prograsp Forceps, LND-Large Needle Driver, VS/SI- Vessel Sealer/ Suction Instrument, GR/CA- Grasp-ing Retractor/Clip Applier, MCS-Monopolar Curved Scissors, UP-Ultrasound Probe)	78
5.4	Performance of SOTA instance segmentation methods on EETS instrument segmentation	79
5.5	Ablation Studies of the proposed S3NET on EV17	79
5.6	Data distribution of classes in the train dataset	82
5.7	Test set results on the challenge IOU metric before post processing.	87
5.8	Test set results on the challenge IOU metric after post processing of SOTA instance segmentation methods.	88
5.9	Performance of SOTA instance segmentation methods on EV17 instrument segmentation dataset. APm represents Mean Average Precision of the mask, i.e., AP0.50:0.95	89
5.10	4-fold cross-validation results on the cumulative IOU metric.	90
6.1	SROCC values on JIGSAWS for the 4-Fold setting	98
6.2	SROCC values on NETS dataset for test set	98

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
PicSO	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