Revolutionizing Gastrointestinal Endoscopy: AI-Driven Surgical Scene Understanding and Video Analysis

QIU YANG[1], ZHU ZHUO[2]
[1]Southern University of Science and Technology, School of Medicine (e-mail: 12111246@mail.sustech.edu.cn)
[2]Southern University of Science and Technology, Department of Biomedical Engineering (e-mail: 12011119@mail.sustech.edu.cn)

This work is the final survey about AI surgical scene understanding in MED331, 2023 Fall.

ABSTRACT Gastrointestinal endoscopy, a critical diagnostic and therapeutic tool in gastroenterology, has been revolutionized by the integration of artificial intelligence (AI). This paper explores AI's role in enhancing the accuracy and efficiency of surgical video analysis in gastrointestinal procedures, with a focus on semantic segmentation and understanding complex surgical scenes. We discuss the advancements achieved through innovative AI models, which have notably improved the segmentation of small or subtle objects within surgical environments. Despite these advancements, the field faces ongoing challenges, including managing variable lighting conditions, motion artifacts, and the necessity for extensive, accurately annotated datasets. The paper also examines potential future developments in surgical AI, such as more sophisticated algorithms, improved data management, and the evolution of autonomous surgical robots. Ethical, regulatory, and global accessibility considerations of these technologies are also addressed. This study highlights the significant impact of AI in gastrointestinal endoscopy, providing insights into its current applications and future possibilities in enhancing surgical practices and patient care.

INDEX TERMS Gastrointestinal endoscopy, artificial intelligence, scene segmentation, surgical video analysis, deep learning algorithms.

I. INTRODUCTION

Gastrointestinal endoscopy is an essential procedure in modern gastroenterology, offering vital insights into the diagnosis and treatment of various digestive disorders. The advent of artificial intelligence (AI) has ushered in a new era in this field, enhancing not only the precision of diagnostic techniques but also the efficacy of therapeutic interventions. This paper aims to elucidate the transformative role of AI in gastrointestinal endoscopy, particularly in the realm of surgical video analysis.

Recently, AI has become a prominent force in medical imaging, greatly enhancing the precision of semantic segmentation and the analysis of intricate surgical scenarios. These developments are crucial for more effective identification and treatment of gastrointestinal conditions. This paper provides a straightforward introduction to this technique and discusses the construction of a scene segmentation model using advanced AI methods, highlighting their impact on the field of gastrointestinal endoscopy.

Moreover, this paper delves into the future prospects of AI in gastrointestinal endoscopy, discussing the potential for advancements in algorithmic development, data management, and the creation of autonomous surgical systems. We also consider the ethical and regulatory implications of incorporating AI into clinical practice, as well as the importance of making these advanced technologies accessible globally. Through this exploration, we aim to provide a comprehensive overview of the current state and future potential of AI in transforming gastrointestinal endoscopy and enhancing patient care in gastroenterology.

II. GASTROINTESTINAL ENDOSCOPY: SOPHISTICATED IMAGING TECHNIQUES

A. OVERVIEW AND MARKET TRENDS IN GASTROINTESTINAL ENDOSCOPY

Gastrointestinal endoscopy is an advanced imaging technology extensively employed for the diagnosis and treatment of digestive system disorders, plays a precise and indispensable role in clinical practice. In the United States alone, there were approximately 15 to 20 million endoscopic examinations conducted in 2016. According to statistical data, the market size of this field reached $30.38 billion by 2022 and is projected to grow further to $32.62 billion in 2023 with a compound annual growth rate of 7.59%. It is speculated that by 2030, the market size will soar to an estimated $54.56 billion [1].

Gastrointestinal endoscopy primarily comprises two components: endoscopy and intraoperative imaging systems. Endoscopy functions as a detection instrument that integrates image sensors, optical lenses, light sources, and mechanical devices. The intraoperative imaging system performs targeted image processing based on the specific characteristics of different surgeries to meet clinical diagnosis and treatment requirements. Gastrointestinal endoscopy enables direct observation of mucosal conditions in the esophagus, stomach, duodenum, colon, and other digestive ducts while detecting abnormal changes such as ulcers, inflammation, hemorrhage, and tumors. Furthermore, it allows for biopsy or surgical interventions such as polyp removal and stent placement through endoscopic procedures [2]. Compared to traditional surgical methods, this technique generally offers advantages including reduced trauma levels, faster recovery times, and enhanced safety measures. However, in most conventional interventional endoscopy procedures, the anatomical targets and positions of surgical tools cannot be observed within the field of vision. Thus, surgeons anticipate obtaining clear and intuitive visualizations of intraoperative images along with accurate real-time localization capabilities for areas of interest and surgical instruments; thus highlighting the significance of developing advanced endoscopic navigation systems.

The top ten topic areas in GI endoscopy were identified by the editorial board of the American Society of Gastrointestinal Endoscopy through an extensive literature search encompassing high-impact medical and gastroenterological journals published in 2022. Remarkably, AI emerged as the foremost area on this esteemed list. AI is being involved in the clinical practice of gastrointestinal endoscopy technology at an alarming rate, and helping to improve a series of endoscopy content such as lesion detection, disease classification, and real-time decision support. The field of gastrointestinal endoscopy technology has witnessed significant advancements in the integration of AI. Researchers have successfully utilized real-time AI to accurately predict clinical recurrence in patients with ulcerative colitis, while also developing deep learning tools capable of precisely describing the location and severity of this disease. Two meta-analytical studies demonstrate that assisted polyp detection using AI significantly enhances diagnostic accuracy and improves lesion detection effectiveness [3].

B. CHALLENGES IN TRADITIONAL ENDOSCOPIC PROCEDURES

Despite the numerous advantages of minimally invasive endoscopic surgery compared to traditional open surgery, such as significantly shorter hospital stays and recovery periods, smaller incisions and scars, lower complication and trauma risks, reduced pain and discomfort, and potentially lower healthcare costs, it also faces several inherent challenges.

Firstly, in terms of perception at the surgical site, due to the lack of depth perception and the complex topology and photometric characteristics of tissues, there may be blind spots and significant visual changes during the surgical process, increasing the complexity of gastrointestinal examination and diagnosis [4]. Therefore, important lesions may be missed or misdiagnosed, and certain tissue areas may be overlooked.

Secondly, regarding the operation of endoscopes and surgical tools, minimally invasive surgery requires precise navigation within deformed and narrow anatomical spaces, coupled with the issue of disparate axes between the hand and eye [5]. This not only demands a high level of expertise but also requires exceptional dexterity in handling surgical instruments.

In addressing these challenges, the application of artificial intelligence technology offers a new perspective and solution. By enabling AI to understand surgical videos, AI not only has the potential to overcome existing limitations and address issues of surgical scene perception and operation but may also lead the future development of surgical procedures.

With the rapid development of artificial intelligence, particularly in the field of computer vision, over the past decade, the next generation of interventional capabilities is likely to be built upon AI modules that can extract information from rich surgical records and provide computer-assisted interventions (CAI) during and after surgery.

III. ARTIFICIAL INTELLIGENCE IN SURGICAL VIDEO COMPREHENSION

All these advanced medical technology concepts are built upon a crucial premise: achieving a deep understanding of semantic information in surgical scenes through artificial intelligence algorithms to effectively address challenges in surgery. Computer vision, as an important branch of artificial intelligence, focuses on enabling computers and systems to extract, analyze, and comprehend complex information from digital images or videos. This comprehensive understanding of visual data allows for the automation and enhancement of various tasks that traditionally require human vision. In the context of gastrointestinal endoscopy and minimally invasive surgery, computer vision algorithms play a pivotal role in computer assist intervention.

A. COMPUTER VISION

Over the past two decades, computer vision technology has experienced rapid development, which enables the aritifical intelligence algorithm to fully understand the world by vision.

Initially, the field of computer vision relied on hand-crafted features [6] and support vector machines (SVM) [7] for image classification. With the advent of the deep learning era, convolutional neural networks (such as AlexNet [8] and ResNet [9]) made significant strides in algorithms. Recently, algorithms based on Transformer models (such as ViT [10] and Swin Transformer [11]) have brought about a new revolution (Figure 1), excelling in various tasks on various datasets (such as ImageNet, COCO [12]) including image classification, object detection, and semantic segmentation, sometimes even surpassing human performance in certain aspects.

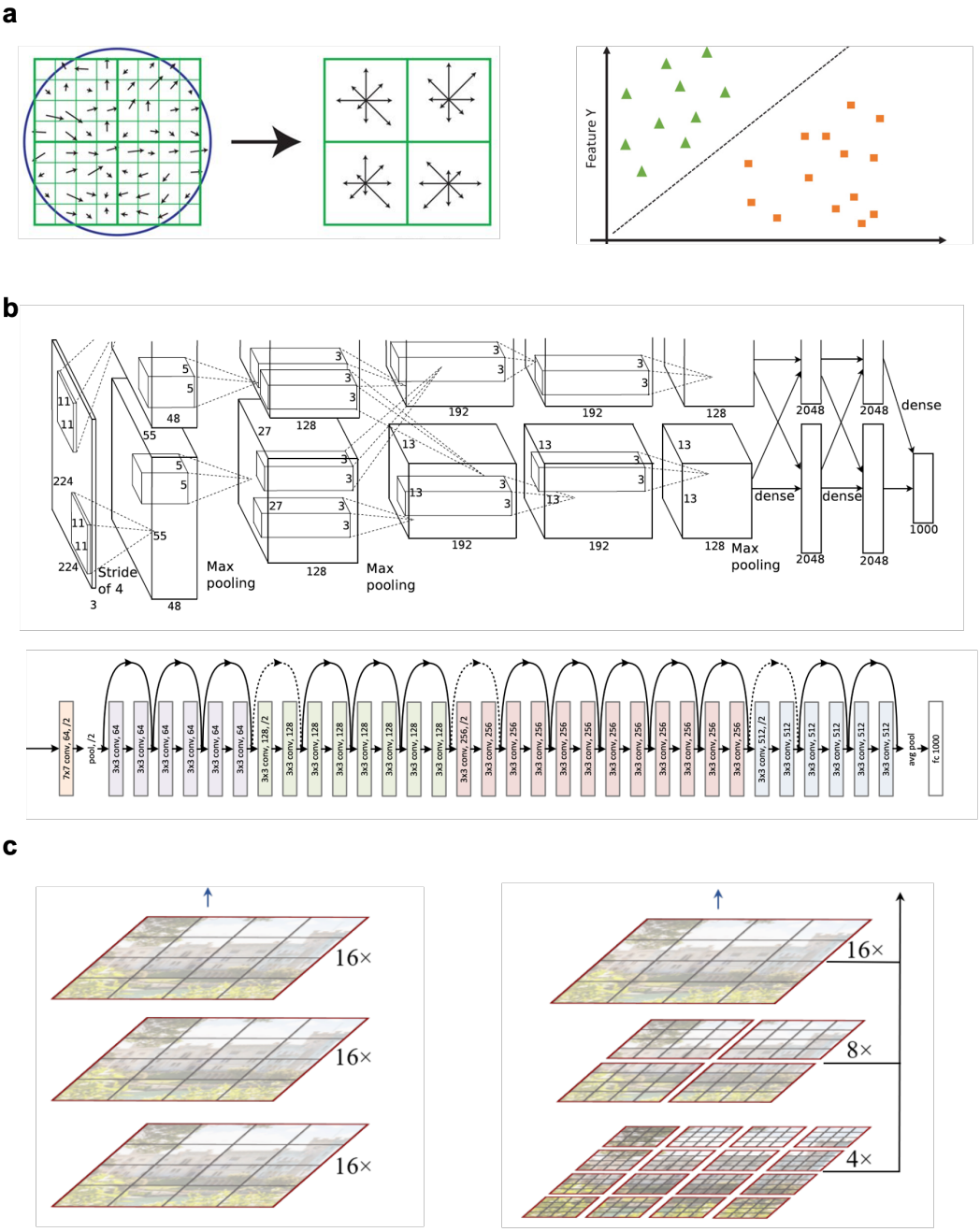

FIGURE 1. The development of vision related algorithm. (a): Illustration of hand-crafted feature extraction using algorithms such as SIFT (Scale-Invariant Feature Transform). This process identifies and describes key points in the image, which are then utilized for image classification through a Support Vector Machine (SVM). (b): Depiction of a Convolutional Neural Network (CNN) architecture, demonstrating the self-learning layers that automatically detect and learn features from input images to solve computer vision tasks. (c): Representation of the Vision Transformer (ViT) model, showcasing the application of attention mechanisms to image processing, which has elevated the capabilities of computer vision algorithms.

In the realm of video understanding, computer vision has also made remarkable advancements in related tasks through innovative improvements to traditional deep learning algorithms, such as the adoption of two-stream networks [13] or three-dimensional video networks [14]. The development of these technologies not only provides powerful tools for analyzing surgical videos but also opens up new possibilities for future medical applications.

B. SURGICAL VISION

In the field of surgical video analysis, a key task for artificial intelligence is to deeply understand the surgical scene both in terms of time and space. In the time dimension, AI needs to identify the specific meaning of each moment in the

surgical process. This understanding can unfold at different levels, ranging from recognizing the overall type of surgery to distinguishing specific actions performed by the surgeon within a certain time frame, further to analyzing specific activities within a few frames of video, and ultimately parsing the specific actions taking place within a single frame. In the spatial dimension, AI must interpret the content within a single frame of the surgical video in detail. This includes understanding from a broader perspective down to precise recognition at different levels: first identifying the organs and surgical instruments present in a single frame, then determining the specific locations of these organs and instruments, followed by semantic segmentation of the organs and surgical instruments, and finally achieving finer instance segmentation of the organs and instruments [15] (Figure 2).

For various surgical scene tasks, there are now many publicly available datasets for AI to learn from, such as those for tool segmentation (EndoVis2018 [16], CholecSeg8k [17], RoboTool [18]), organ segmentation (Dresden Surgical Anatomy Dataset [19], SurgAI3.8K [20]), tool-tissue action detection (CholecT50 [21], SARAS-MESAD2021, PSI-AVA [23]), and skill assessment and workflow recognition (JIGSAWS [24], HeiCO [25], MISAW [26]).

Through this kind of video understanding, artificial intelligence provides strong support for more precise and efficient surgical assistance. This not only enhances the quality of decision-making during surgery but also lays the foundation for the development of future automated surgical technologies.

C. STSWIN IN SEGMENTATION TASK

In our final course project, we developed an artificial intelligence model for understanding surgical videos, based on the STswin Transformer [27]. This model innovatively employs the Swin Transformer across both spatial and temporal dimensions, marking a significant advancement over traditional models that rely on CNN-LSTM aggregation modules. The STswin Transformer's unique space-time window shift mechanism enables efficient processing of spatial and temporal information, leading to a more detailed pixel-level analysis (Figure 3). This methodology effectively tackles common issues in surgical video analysis, such as indistinct decision boundaries and class imbalances. Additionally, the model's adaptable structure makes it suitable for a variety of network configurations and applicable in scenarios where understanding temporal aspects is crucial.

In our project, we implemented Dice loss for scene segmentation using CholecSeg8k dataset. The model demonstrates a certain level of understanding of the surgical environment. However, there remains significant room for improvement. It struggles with segmenting certain elements, particularly small and thin objects like graspers, as well as connective tissue, which tends to be less distinct compared to other types of tissue.

IV. AI IN SURGICAL VISION: DOWNSTREAM TASKS

The realm of modern endoscopic surgery is witnessing a revolutionary transformation through the integration of cutting-edge technologies. These advancements, primarily driven by artificial intelligence (AI) and robotics, are redefining the landscape of minimally invasive procedures. With the realm that artificial intelligence fully understand the surgical scene by vision related algorithms, the implementation of the systems is not just an innovation but a necessity in the contemporary surgical setting. This section delves into the pivotal developments and applications of these technologies, including computer-assisted detection and diagnosis, endoscopic mapping, the emergence of fully automatic surgical robots, surgical training, and automatic generation of electronic medical records (Figure 4).Each of these components represents a significant stride in the journey towards more accurate, efficient, and patient-centric surgical care.

A. COMPUTER-ASSISTED DETECTION & DIAGNOSIS

In endoscopy, computer-aided detection (CADe) and computer-aided diagnosis (CADx) solutions are primarily employed for the identification and classification of abnormal tissue regions. Through learning and training, computers can automatically recognize abnormal tissues and provide diagnostic assistance. Early CAD methods relied on manually designed features, whereas modern deep learning techniques leverage the intrinsic features inherent in the data itself for learning, without assuming specific appearances or texture patterns of diseases a priority. This enables a more accurate and robust classification of abnormal tissues. CADe and CADx systems have been successfully implemented in various upper gastrointestinal endoscopic examinations, significantly enhancing lesion detection rates and diagnostic accuracy, particularly for colorectal tumors. Studies utilizing state-of-the-art convolutional neural networks (CNNs) based on NBI technology have demonstrated that CADx methods can effectively differentiate between five different types of colorectal lesions with a precision detection rate exceeding 95% for colorectal polyps [28].

B. ENDOSCOPIC MAPPING, ANATOMICAL STRUCTURE RECOGNITION AND NAVIGATION

Conventional endoscopy simultaneous localization and mapping (SLAM) approaches rely on complex photogrammetry pipelines to infer the geometry and endoscopic displacement of the gastrointestinal tract while capturing a series of endoscopic images. However, due to the intricate topological characteristics of the gastrointestinal tract and Clinical emergencies, such as hemorrhage, detecting and tracking visual features pose challenges in endoscopy [29]. The incorporation of artificial intelligence has instigated a transformative shift in various aspects of endoscopic navigation: Firstly, CNN-based SLAM techniques enable direct estimation of depth maps from a single monocular view, thereby obviating the necessity for visual feature tracking. In comparison to conventional approaches, these methods exhibit precise

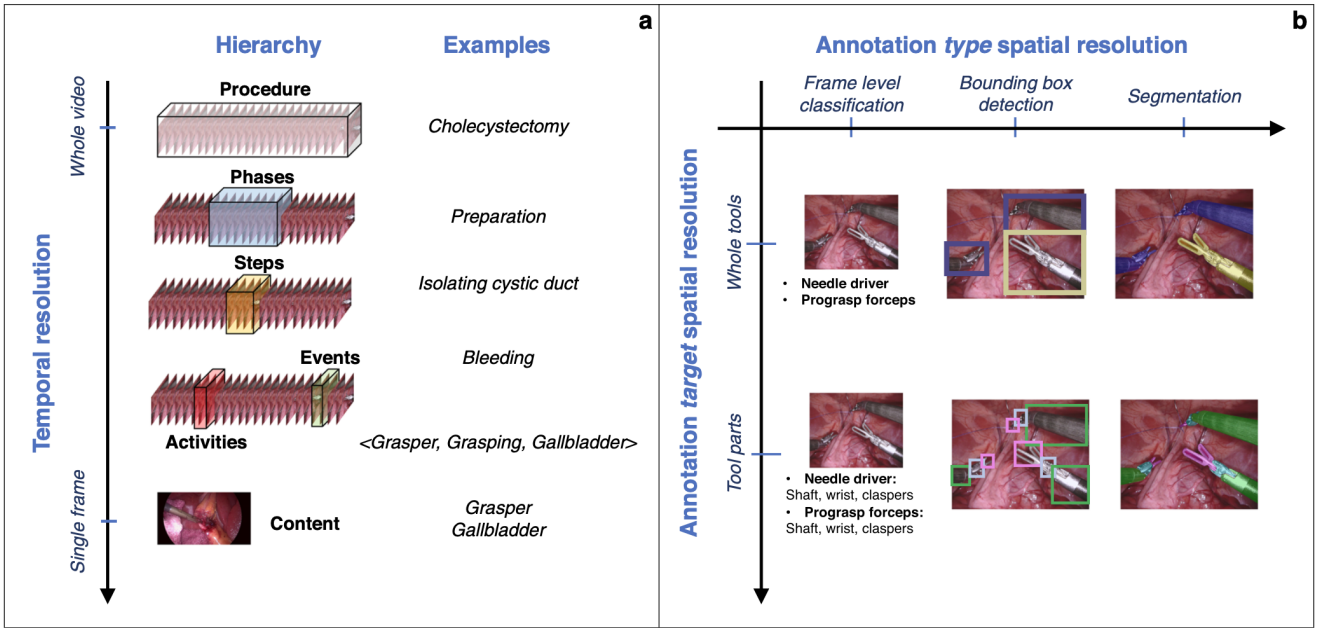

FIGURE 2. Framework for the analysis of endoscopic videos. Temporal (a) and spatial (b) annotations at different resolutions are used to model tasks at increasingly finer details [15].

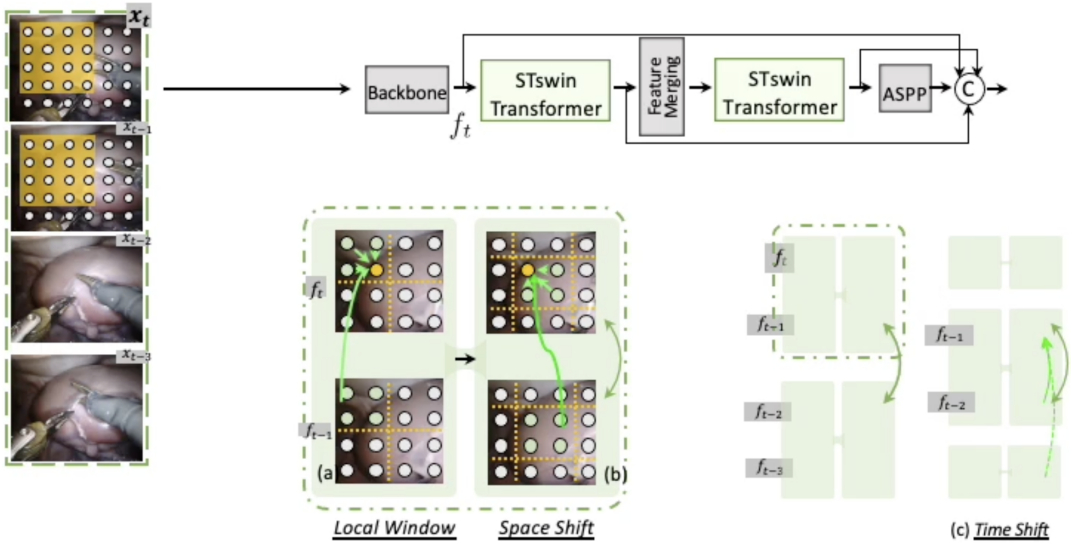

FIGURE 3. This diagram of the STswin architecture: STswin employs a shifting window mechanism across both spatial and temporal dimensions for feature extraction. The model processes a sequence of four consecutive frames, effectively capturing and analyzing features from adjacent frames within the series [27].

mapping and positioning outcomes on extended colonoscopy sequences, showcasing their efficacy in delivering accurate results [30]. Secondly, Artificial intelligence technology facilitates precise recognition of anatomical structures during endoscopic navigation, providing enhanced assistance. The goal of anatomical structure recognition is to identify different segments of the digestive tract as well as key structures or landmarks. Studies have demonstrated that traditional convolutional neural network (CNN) architectures achieve classification accuracy rates exceeding 85% for these structures [31]. Moreover, artificial intelligence can be leveraged to develop innovative robotic platforms and paradigms such as magnetic endoscopes, which offer improved navigation support.

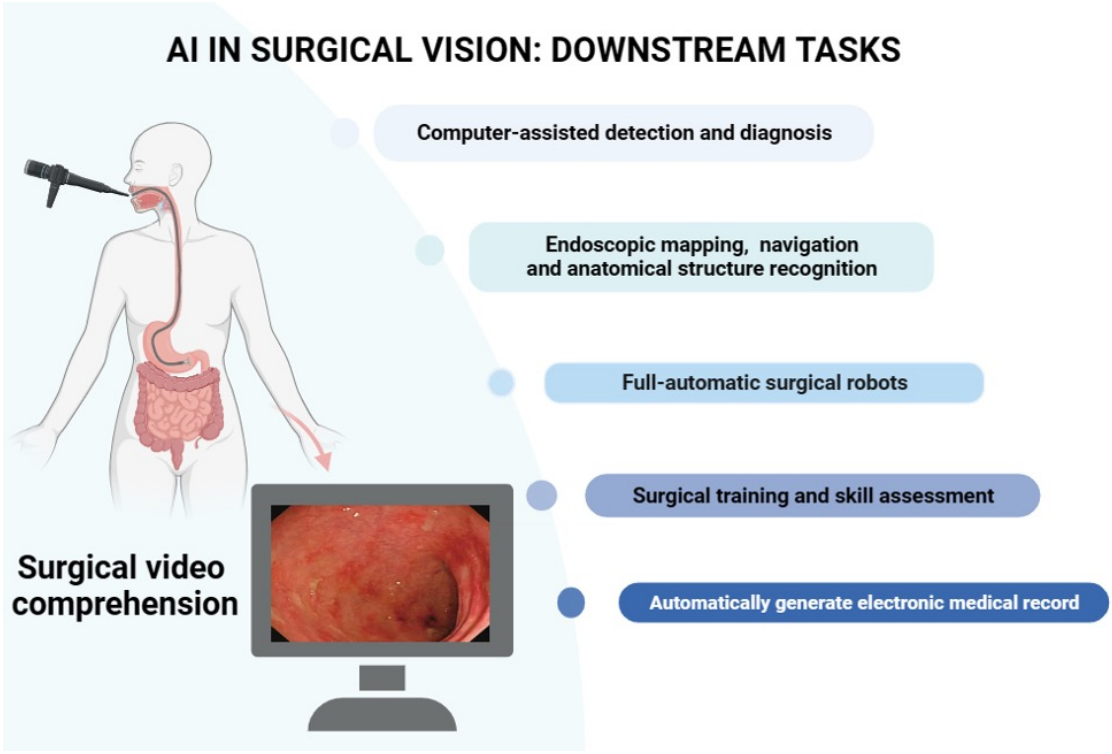

FIGURE 4. The infographic presenting the downstream tasks facilitated by advancements in surgical scene understanding. It illustrates the integration of computer-assisted technologies in various aspects of endoscopic surgery, including detection and diagnosis, endoscopic mapping, navigation, anatomical structure recognition, and the operation of full-automatic surgical robots. Additional applications highlighted include the use of AI for surgical training and skill assessment, as well as the automatic generation of electronic medical records, demonstrating the comprehensive impact of surgical video comprehension on enhancing and streamlining gastrointestinal endoscopy procedures.

C. FULL-AUTOMATIC SURGICAL ROBOTS

Automatic surgical robots are robotic systems capable of performing surgeries without human intervention, with a primary focus on autonomy and intelligence. Autonomy research emphasizes the independent decision-making and operational abilities of robots during surgery, encompassing path planning, tissue recognition, and execution. Intelligence research centers around the perception and cognitive capabilities of robots, including image processing, speech recognition, and machine learning. Currently, certain studies have achieved autonomous decision-making and operational abilities in areas such as path planning and tissue recognition. Moreover, advancements in robot intelligence have been made through technologies like image processing and machine learning applications. Automatic surgical robots possess immense potential to enhance surgical procedures through further research aimed at attaining higher levels of autonomy and intelligence.

D. SURGICAL TRAINING AND SKILL ASSESSMENT

The assessment of surgical skills is employed to evaluate the professional proficiency of surgeons during specific surgical tasks. Currently, automated evaluation employing temporal neural networks is the predominant method for surgical skill assessment. These networks have been trained on the JIGSAWS dataset to differentiate between three levels of physician expertise while identifying surgical gestures and assessing skill scores in robot-assisted urology with an accuracy exceeding 95%. However, their effectiveness and impact are limited by the absence of real clinical data. Additionally, a three-stage temporal neural network approach has been developed for laparoscopic cholecystectomy achieving an average classification accuracy of approximately 85% on a proprietary cholecystectomy dataset for distinguishing between proficient and inadequate surgical skills. Nevertheless, extensive clinical datasets are necessary to validate this method's reliability.

E. AUTOMATICALLY GENERATE ELECTRONIC MEDICAL RECORDS

By harnessing advanced technologies such as machine learning and image processing, the automatic recording and analysis of surgical procedures can be achieved through video data acquired from endoscopic surgery. These records encompass both functional and structural information about the patient's anatomy, accompanied by a comprehensive log documenting events, activities, and procedures throughout the surgical intervention. These comprehensive records can furnish clinicians with meticulous information, facilitating more precise diagnosis and treatment decisions.

V. FUTURE OF SURGICAL AI

A. LIMITATIONS AND CHALLENGES

Video is the most common data type in surgical phase recognition. The current technological limitations hinder the implementation of endoscopic video traceability, which effectively safeguards patient privacy and provides extensive research opportunities for academic groups. However, irregular scenes present a significant challenge to endoscopic video applications. Frames captured at the same stage may exhibit substantial variations due to these irregular scenes. Sudden movements of the endoscopic camera during procedures can result in displacement and dislocation. Moreover, changes in lighting conditions, blood, artifacts resulting from lens cleaning processes, and other factors can adversely impact image quality.

The acquisition of surgical video comprehension heavily relies on the availability of large-scale annotated datasets. Despite notable advancements in machine learning techniques within gastrointestinal endoscopy, challenges persist due to limited representative data labels and inconsistent data quality. The performance of a proposed model trained using a restricted dataset may be significantly affected when applied to test datasets from unfamiliar medical sources. Endoscopic images exhibit inherent heterogeneity, and their high dimensionality and volume can substantially impact the efficacy of computer vision systems [32]. In certain cases, there might be an insufficient number of training samples in the dataset to effectively train deeper or wider networks. Although EndoVis, Cholec80, and JIGSAWS are currently recognized as widely used public datasets, there remains a need for larger and accurately labeled datasets to be established.

With the increasing use of artificial intelligence in healthcare, ensuring the protection and confidentiality of patient data has become a pressing concern. It is essential to establish clear definitions of responsibility and legal liability for AI systems used in medical diagnosis, treatment, and prediction. Currently, there are gaps in laws and regulations pertaining to AI assistance and automation, necessitating comprehensive consideration and an appropriate solution.

B. FUTURE DEVELOPMENT

The future of surgical artificial intelligence is poised to be marked by significant advancements that will transform the landscape of surgery and healthcare. Key developments will focus on enhancing the precision and efficiency of AI algorithms, particularly in the realm of computer vision for more accurate analysis of surgical videos. There will be a push towards creating larger, diverse, and accurately labeled datasets to improve the training and effectiveness of AI systems.

Robotics will play an increasingly pivotal role, with developments in autonomous surgical robots that integrate advanced AI for more precise and less invasive procedures. Ethical and regulatory frameworks will also evolve, addressing the responsibilities and liabilities associated with AI in surgery and ensuring patient data privacy and security.

Additionally, AI's role in surgical education and training will expand, utilizing technologies like virtual and augmented reality for immersive and realistic training experiences. Lastly, efforts will be made to ensure global accessibility of these advancements, making cutting-edge surgical AI tools available in diverse healthcare settings, thereby democratizing high-quality surgical care.

VI. DISCUSSION AND CONCLUSION

The exploration and integration of artificial intelligence in surgical applications, particularly in gastrointestinal endoscopy, have opened new frontiers in medical technology. The incorporation of AI has shown immense potential in enhancing the accuracy and efficiency of surgical procedures.

One of the key aspects of this integration is the improvement in semantic segmentation and the ability to understand complex surgical scenes. Innovative AI models improved the segmentation of subtle or small objects, such as surgical instruments and various tissue types. Despite these advancements, broader challenges persist in the AI domain. These include coping with variable lighting conditions, motion artifacts in video data, and the pressing need for extensive, precisely annotated datasets. These hurdles highlight the ongoing journey of AI development in accurately interpreting complex surgical environments.

Looking ahead, the development of surgical AI promises advancements in algorithmic sophistication, data management, and enhanced computer vision systems. The evolution of autonomous surgical robots and the establishment of comprehensive regulatory and ethical frameworks are critical areas of focus. Additionally, AI's expanding role in surgical training and the push for global accessibility underscore the transformative impact of these technologies.

In conclusion, the integration of AI in surgical applications marks a pivotal shift in healthcare, offering unprecedented precision and efficiency in surgical procedures. The future holds promising advancements that are likely to revolutionize surgical practices. However, this progress must be balanced with ethical considerations, continuous learning, and adaptation to new challenges and discoveries. As the field of surgical AI continues to evolve, it will undoubtedly play a crucial role in shaping the future of surgery and patient care.

REFERENCES

[1] Mariam Faizullabhoy , Shishanka Wangnoo, "Endoscopy Market - By Product (Endoscopes, Visualization Systems, Endoscopic Ultrasound, Insufflator), By Application (Arthroscopy, Laparoscopy, GI Endoscopy, Obstetrics/Gynecology, ENT Endoscopy, Pulmonary Endoscopy), By End-use & Forecast, 2023 – 2032", Aug 2023.

[2] Luo, X., Mori, K., & Peters, T. M. (2018). Advanced Endoscopic Navigation: Surgical Big Data, Methodology, and Applications. Annual review of biomedical engineering, 20, 221–251.

[3] Mulki, R., Qayed, E., Yang, D., Chua, T. Y., Singh, A., Yu, J. X., Bartel, M. J., Tadros, M. S., Villa, E. C., & Lightdale, J. R. (2023). The 2022 top 10 list of endoscopy topics in medical publishing: an annual review by the American Society for Gastrointestinal Endoscopy Editorial Board. Gastrointestinal endoscopy, 98(6).

[4] Bogdanova,R.,Boulanger,P. &Zheng,B.Depth perception of surgeons in minimally invasive surgery. Surg. Innov. 23, 515–524 (2016).
[5] Martin, J. W. et al. Enabling the future of colonoscopy with intelligent and autonomous magneticmanipulation.Nat.Mach.Intell.2,595–606(2020).
[6] Lindeberg, Tony. "Scale invariant feature transform." (2012): 10491.
[7] Hearst M A, Dumais S T, Osuna E, et al. Support vector machines[J]. IEEE Intelligent Systems and their applications, 1998, 13(4): 18-28.
[8] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." Advances in neural information processing systems 25 (2012).
[9] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.
[10] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T. & Houlsby, N. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929.
[11] Liu Z, Lin Y, Cao Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows. Proceedings of the IEEE/CVF international conference on computer vision. 2021: 10012-10022.
[12] Lin T Y, Maire M, Belongie S, et al. Microsoft coco: Common objects in context[C]//Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13. Springer International Publishing, 2014: 740-755.
[13] Simonyan K, Zisserman A. Two-stream convolutional networks for action recognition in videos[J]. Advances in neural information processing systems, 2014, 27.
[14] Carreira J, Zisserman A. Quo vadis, action recognition? a new model and the kinetics dataset[C]//proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 6299-6308.
[15] Mascagni, P., Alapatt, D., Sestini, L., Altieri, M. S., Madani, A., Watanabe, Y. & Hashimoto, D. A. (2022). Computer vision in surgery: from potential to clinical value. npj Digital Medicine, 5(1), 163.
[16] Allan, M., Kondo, S., Bodenstedt, S., Leger, S., Kadkhodamohammadi, R., Luengo, I. & Speidel, S. (2020). 2018 robotic scene segmentation challenge. arXiv preprint arXiv:2001.11190.
[17] Hong, W. Y., Kao, C. L., Kuo, Y. H., Wang, J. R., Chang, W. L., & Shih, C. S. (2020). Cholecseg8k: a semantic segmentation dataset for laparoscopic cholecystectomy based on cholec80. arXiv preprint arXiv:2012.12453.
[18] Garcia-Peraza-Herrera, L. C., Fidon, L., D'Ettorre, C., Stoyanov, D., Vercauteren, T., & Ourselin, S. (2021). Image compositing for segmentation of surgical tools without manual annotations. IEEE transactions on medical imaging, 40(5), 1450-1460.
[19] Carstens, M., Rinner, F. M., Bodenstedt, S., Jenke, A. C., Weitz, J., Distler, M. & Kolbinger, F. R. (2023). The Dresden Surgical Anatomy Dataset for abdominal organ segmentation in surgical data science. Scientific Data, 10(1), 3.
[20] Zadeh S M, François T, Comptour A, et al. SurgAI3. 8K: A Labeled Dataset of Gynecologic Organs in Laparoscopy with Application to Automatic Augmented Reality Surgical Guidance[J]. Journal of Minimally Invasive Gynecology, 2023, 30(5): 397-405.
[21] Nwoye, C. I., Yu, T., Gonzalez, C., Seeliger, B., Mascagni, P., Mutter, D. & Padoy, N. (2022). Rendezvous: Attention mechanisms for the recognition of surgical action triplets in endoscopic videos. Medical Image Analysis, 78, 102433.
[22] D. Ebehard and E. Voges, "Digital single sideband detection for interferometric sensors," presented at the 2nd Int. Conf. Optical Fiber Sensors, Stuttgart, Germany, Jan. 2-5, 1984.
[23] Valderrama, N., Ruiz Puentes, P., Hernández, I., Ayobi, N., Verlyck, M., Santander, J. & Arbeláez, P. (2022, September). Towards holistic surgical scene understanding. In International conference on medical image computing and computer-assisted intervention (pp. 442-452).
[24] Gao, Y., Vedula, S. S., Reiley, C. E., Ahmidi, N., Varadarajan, B., Lin, H. C., & Hager, G. D. (2014, September). Jhu-isi gesture and skill assessment working set (jigsaws): A surgical activity dataset for human motion modeling. In MICCAI workshop: M2cai (Vol. 3, No. 3).
[25] Maier-Hein, L., Wagner, M., Ross, T., Reinke, A., Bodenstedt, S., Full, P. M. & Müller-Stich, B. P. (2021). Heidelberg colorectal data set for surgical data science in the sensor operating room. Scientific data, 8(1), 101.
[26] Huaulmé, A., Sarikaya, D., Le Mut, K., Despinoy, F., Long, Y., Dou, Q. & Jannin, P. (2021). Micro-surgical anastomose workflow recognition challenge report. Computer Methods and Programs in Biomedicine, 212, 106452.
[27] Jin, Y., Yu, Y., Chen, C., Zhao, Z., Heng, P. A., & Stoyanov, D. (2022). Exploring intra-and inter-video relation for surgical semantic scene segmentation. IEEE Transactions on Medical Imaging, 41(11), 2991-3002.
[28] Chadebecq, F., Lovat, L. B., & Stoyanov, D. (2023). Artificial intelligence and automation in endoscopy and surgery. Nature reviews. Gastroenterology & hepatology, 20(3), 171–182.
[29] Demir, Kubilay Can; Schieber, Hannah; Roth, Daniel; Maier, Andreas; Yang, Seung Hee (2022). Surgical Phase Recognition: A Review and Evaluation of Current Approaches. TechRxiv. Preprint.
[30] Ma, R. et al. RNNSLAM: reconstructing the 3D colon to visualize missing regions during a colonoscopy. Med. Image Anal. 72, 102100 (2021).
[31] Mahmoud, N. et al. Live tracking and dense reconstruction for handheld monocular endoscopy. IEEE Trans. Med. Imaging 38, 79–89 (2019)
[32] Bayoudh, K., Knani, R., Hamdaoui, F. et al. A survey on deep multimodal learning for computer vision: advances, trends, applications, and datasets. Vis Comput 38, 2939–2970 (2022).