Advancing Automation in Microscopic and Endoscopic Neurosurgical Skills Assessment

Abstract

Minimally invasive neurosurgery demands high precision in tasks like micro-suturing, micro-drilling, and endoscopic navigation. Due to patient safety concerns and legal restrictions, surgical training has shifted from operating rooms to simulation labs. While several training systems exist, skill evaluation still depends heavily on manual expert review, which is time-consuming and inconsistent. This thesis proposes automated skill evaluation methods for both microscopic and endoscopic procedures. We introduce image-based solutions for microscopic skills evaluation and a video-based solution for endoscopic skills evaluation.

For micro-suturing, we introduce a multi-task learning model based on modified Res2Net, which predicts both an overall score and five detailed sub-scores, achieving a 30.26% improvement over existing methods. A novel logit adjustment for regression further enhances performance by 4.75%. For micro-drilling, where data is limited, we formulate the problem into a rank estimation framework using geometric order learning and demonstrate that the proposed formulation outperforms the state-of-the-art classification and regression techniques. In endoscopic tasks, we design a two-stage pipeline. The first stage involves temporal action segmentation (TAS) to identify specific actions in the video. In the second stage, each action is evaluated. We propose a new data augmentation strategy tailored for video sequences that maintains optical flow consistency and generates pseudo action boundaries to improve the performance in data-scarce conditions by around 7.89% in edit scores. On average, it outperforms the best existing data augmentation method by 3.94%. Further, in the second stage, a multimodal contrastive pretraining framework is also proposed to create semantically meaningful video embeddings using textual skill descriptions, better suited for low-data settings in order to achieve accurate and reliable skills evaluation.

To bridge simulation and real surgery, we improve surgical instrument instance segmentation by proposing a Query Proposal Network with deformable attention, improving query recall by 44.89% and segmentation accuracy by over 2% on standard benchmarks. Alongside algorithms, this thesis contributes four large-scale, annotated datasets across micro-suturing, micro-drilling, and endoscopic skills, including the largest open-access dataset for micro-suturing and the first temporal action dataset for non-robotic neuro-endoscopy. Real-world deployment studies show strong alignment between expert and automated scores, with an MSE of 0.62 on a Likert scale of 10 and high user satisfaction (4.32/5), demonstrating the system's reliability and practicality. Overall, this work provides a unified, data-driven framework for objective, scalable, and real-world-compatible neurosurgical skill evaluation.