Open

Citation: J Med Discov (2026);11(2):jmd26008; DOI:10.24262/jmd.11.2.26008

Research Article

Deep Learning-Based Anti-Collision System for Endoscopes in Minimally Invasive Spinal Surgery

Xiaowei Song[1]

[1]*Maoyu (Qingdao) Medical Technology Co., Ltd. Beijing Branch. Address:No. 156, Beiqing Road, Beijing Zhongguancun Environmental Protection Park Innovation Center,Haidian District, Beijing,100080, China.*

Abstract Unintentional collisions between surgical knife tips and endoscopes are a primary cause of damage to expensive medical equipment in minimally invasive spinal surgery (MISS). A single endoscope costs approximately 15,000 USD, with an average maintenance cost of several thousand USD. Such collisions also lead to surgical interruptions and an increased risk of complications. To address this clinical challenge, this paper proposes a two-stage deep learning-based endoscopic anti-collision algorithm: the YOLOv8 model is used for accurate detection and existence confirmation of knife tips within the surgical field of view, and the MobileNetV3-Large model is integrated for binary classification of the dangerous distance state between knife tips and endoscopes. The system is developed on a GPU-accelerated hardware platform, incorporating a CUDA-optimized image preprocessing pipeline and a real-time visual feedback module, and has been successfully deployed on a 4K@60fps surgical endoscopic camera system. During algorithm development and model training, two dedicated datasets were constructed for surgical knife tip detection and knife tip danger state classification, respectively. Experimental validation shows that the knife tip detection achieves a mAP@50 of 94.6%, the danger state classification accuracy reaches 92.29%, and the overall processing frame rate of the system stabilizes at 60 FPS, meeting the real-time clinical requirements. Clinical application results demonstrate that the system can reduce the endoscopic collision damage rate to zero and cut the annual average equipment maintenance cost by about 800 USD per rigid endoscope, exhibiting significant engineering practical value and clinical economic benefits.

Keywords: Minimally invasive spinal surgery; Endoscope protection; Knife tip detection; Distance perception; Collision protection; YOLOv8; MobileNetV3

How to cite: Xiaowei Song et al., Deep Learning-Based Anti-Collision System for Endoscopes in Minimally Invasive Spinal Surgery. J Med Discov (2026); 11(2):jmd26008; DOI:10.24262/jmd.11.2.26008; Received February 06th, 2026, Revised March 19th, 2026, Accepted March 31st, 2026, Published April 06th, 2026.

1.Introduction

1.1 Engineering Background and Problem Statement

Minimally invasive spinal surgery (MISS) has become the mainstream treatment for spinal diseases due to its advantages of minimal trauma, less bleeding and rapid recovery. However, the narrow surgical space and limited depth perception lead to a high risk of collisions between surgical knife tips and the lenses of rigid endoscopes [1][2]. As the core carrier of the surgical field of view, the optical lens assembly of an endoscope is manufactured with high-precision processes, with a single device generally costing 15,000 to 30,000 USD. Clinical data from the Affiliated Hospital of Xuzhou Medical University shows that 95 rigid endoscopes underwent 103 maintenance procedures from January to December 2022, 22% of which were due to component damage, with the total maintenance cost reaching 80,000 USD (excluding free maintenance and warranty repairs) [3]. Meanwhile, MISS features a steep learning curve for surgeons and high training difficulty; novice surgeons struggle to judge the actual distance

*Correspondence: Xiaowei Song, Maoyu (Qingdao) Medical Technology Co., Ltd. Beijing Branch. Address: No. 156, Beiqing Road, Beijing Zhongguancun Environmental Protection Park Innovation Center,Haidian District, Beijing,100080, China. Email: xiaowei_song123@163.com.

between the knife tip and the endoscope independently, and there are significant differences in surgical skills among practicing physicians [4].

Existing protective measures rely mainly on surgeons' empirical judgment, lacking an objective distance quantification and active anti-collision protection mechanism. Traditional mechanical protective devices restrict the flexibility of surgical operations, while distance measurement technologies based on ultrasound or infrared suffer from complex hardware integration and insufficient real-time performance. From an engineering application perspective, there is an urgent need to develop a non-invasive, high-precision and real-time responsive anti-collision protection system that can realize distance monitoring and danger early warning between the knife tip and the endoscope without affecting the existing surgical process, thereby reducing the risk of equipment damage from an engineering perspective.

1.2 Research Status of Related Technologies

In the field of surgical instrument detection, deep learning has become the mainstream technical solution. Jin et al. [5] realized laparoscopic instrument localization based on Region-Based Convolutional Neural Networks (R-CNN) with an average precision of 81.8%; Choi et al. [6] adopted the YOLO algorithm for surgical instrument detection, whose real-time performance meets clinical requirements. However, existing detection algorithms mostly focus on the overall localization of instruments, lacking a precise recognition and existence verification mechanism for the knife tip area, which is prone to misjudgment due to background interference.

In terms of danger state recognition, Hasan et al. [7] proposed the ART-Net network, which realizes 3D pose estimation of surgical instruments through geometric primitive extraction with a position measurement error of 2.5 mm; Namazi et al. [8] designed LapTool-Net, which improves the detection robustness in complex scenarios through a context-aware model. Nevertheless, these methods lack engineering optimization for close-range measurement between the knife tip and the endoscope. In terms of engineering implementation, TensorRT and CUDA acceleration technologies have been widely applied in real-time medical image processing [9]. However, how to effectively integrate deep learning algorithms with 4K high-frame-rate endoscopic systems and solve engineering problems such as data transmission delay, excessive hardware resource occupation and clinical deployment compatibility remains a weak link in current research.

1.3 Research Objectives and Core Contributions

Aiming at clinical practical needs and pain points such as the steep learning curve of minimally invasive surgery, this paper designs and implements a directly deployable endoscopic anti-collision protection system. The core objectives are as follows:

(1) Construct datasets for surgical knife tip detection and surgical knife tip danger state classification;

(2) Realize accurate detection and existence confirmation of knife tips to reduce the false alarm rate;

(3) Complete binary classification of danger states based on distance thresholds to ensure early warning accuracy;

(4) Control the start and stop of power and plasma equipment according to the early warning state of the knife tip from the camera system to realize the self-protection mechanism of the endoscope;

(5) Adapt to 4K@60fps endoscopic systems to meet real-time requirements.

The main contributions are as follows:

(1) Two datasets for knife tip detection and danger state classification are constructed, including 2 types of grinding knife tips, 2 types of planing knife tips, 1 type of radio frequency (RF) knife tip and 1 type of plasma knife tip. All knife tips support both UBE and plus working modes, covering surgical and normal scenarios, which ensures the precision and recall of detection and classification;

(2) A detection-classification two-stage early warning architecture is proposed: YOLOv8 is used for accurate knife tip localization and existence verification, and MobileNetV3-Large for danger state classification, balancing detection accuracy and real-time performance;

(3) Based on the classification of knife tip danger states, the camera system controls the start and stop of power and plasma equipment through serial communication, thereby realizing an endoscopic anti-collision protection mechanism in clinical environments;

(4) A GPU-accelerated engineering pipeline is built, integrating CUDA-optimized modules for image preprocessing, model inference and post-processing to ensure real-time processing of 4K video streams;

(5) Engineering deployment and clinical verification of the system are completed to ensure its stable operation in clinical environments.

2. System Design and Implementation

2.1 Overall System Architecture

The system operation diagram is shown in Figure 1, with the main equipment including a camera host, a display, a power host (plasma host), a camera handle, an endoscope and a power knife tip (plasma electrode). The camera host is used to collect video streams and complete a series of processes such as knife tip detection and danger state classification; the power host (plasma host) is controlled by the camera host, which sends start/stop signals to the power host through a serial port according to the danger state. The display is used to show post-processed images, with the danger state label displayed in the upper left corner of the image for the surgeon to grasp the current state in real time.

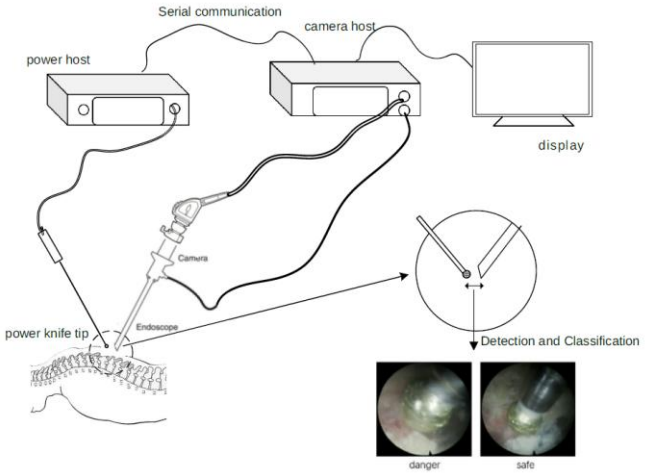

Figure 1 System Work Diagram

The system adopts a modular design concept, and its overall architecture is divided into five core modules: image acquisition and preprocessing module, knife tip detection module, danger state classification module, image post-processing module and system control module. The modules communicate in a pipeline manner to ensure system simplicity (see Figure 2 for the system architecture).

Figure 2 System Architecture

(1) The 4K endoscopic camera collects UYVY format video streams, and completes RGB format conversion, cropping and scaling preprocessing through a CUDA-accelerated pipeline;

(2) The pre-trained YOLOv8 model detects the position of the knife tip and confirms its existence in the surgical field of view, outputting a flag indicating the presence or absence of the knife tip in the field of view;

(3) The pre-trained MobileNetV3-Large model performs binary classification of danger states;

(4) The image post-processing module generates visual early warning markers (safe/danger) according to the classification results and overlays them on the real-time system images;

(5) The system control module is responsible for resource scheduling and linkage control of external equipment.

The system implements a two-stage anti-collision protection mechanism (see Figure 3): preprocessed images are first input to the YOLOv8 model for knife tip detection; if a knife tip is detected, the MobileNet model is then used for knife tip danger state classification. Engineering practice verifies that the complex intraoperative environment (e.g., tissue reflection, bone chip occlusion) can easily lead to misclassification when preprocessed images are directly used for knife tip danger state classification; thus, a knife tip detection model is added to generate a knife tip existence flag. YOLOv8 features lightweight design, high detection accuracy and low deployment difficulty; engineering verification shows that the addition of YOLOv8 detection has a minimal impact on the delay of the actual system and significantly improves the accuracy of knife tip danger state classification.

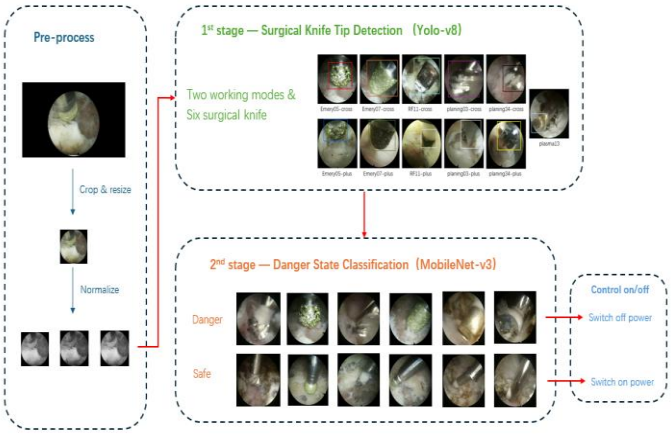

Figure 3 2-stage anti-collision protection mechanism

2.2 Dataset Collection

To train the YOLOv8 knife tip detection model and the MobileNetV3-Large knife tip danger state classification model, two corresponding datasets were constructed in this paper. The datasets support two surgical modes: UBE (a dual-channel mode with a cross layout of endoscope and knife tip) and plus (a single-channel mode with a coaxial layout of endoscope and knife tip). A total of six types of knife tips are included: 2 types of grinding knife tips, 2 types of planing knife tips, 1 type of RF knife tip and 1 type of plasma knife tip.

(1) YOLOv8 knife tip detection dataset: 20 videos of minimally invasive spinal surgery were collected, and corresponding videos were captured in the laboratory under different background conditions. A total of 15,000 frames of images were annotated, of which 80% were used for training and 20% for testing (see Figure 4 for the detailed classification of knife tips). Plasma13 is only applicable to the plus surgical mode, so no cross/plus distinction is made for this type.

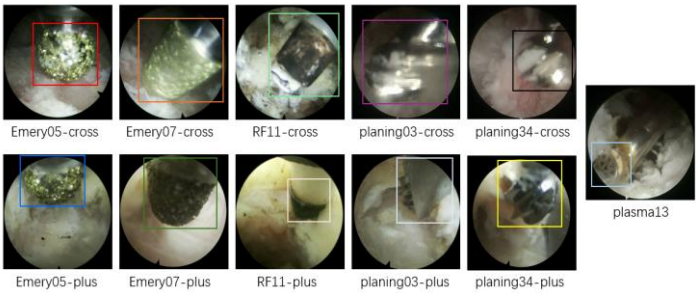

Figure 4 knife tip classification

(2) MobileNetV3-Large knife tip danger state classification dataset: Images of knife tip-endoscope at different distances were collected on a surgical simulation platform, with the dangerous distance threshold calibrated at 3 mm. A classification dataset with 30,000 samples was constructed, with a 1:1 ratio of dangerous to safe samples (see Figure 5 for the classification of knife tip danger states).

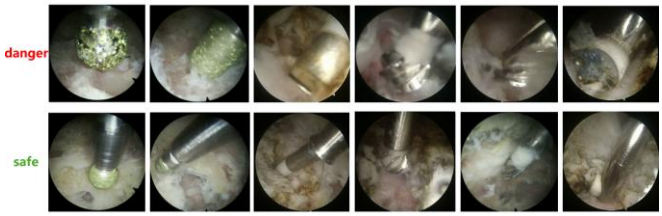

Figure 5 danger state classification

2.3 Hardware Platform Configuration

The system is deployed on the NVIDIA AGX ORIN (32G) hardware platform, whose hardware selection balances real-time performance and hardware volume (see Table 1 for specific configurations). The platform supports CUDA 12.2 and TensorRT 8.6 acceleration, meeting the real-time processing requirements of 4K@60fps video streams, and features excellent heat dissipation and stability.

2.4 Implementation of Core Algorithms

Aiming at the characteristics of endoscopic videos and the processing requirements of 4K resolution, with the core engineering optimizations as follows:

(1) Preprocessing: The CUDA-accelerated kernel functions are used to realize fast conversion from UYVY to RGB format, image downsampling and scaling, ensuring that the images are fed into the model for inference at a resolution of 240x240;

(2) Model inference: The pre-trained models for knife tip detection and danger state classification are converted into TensorRT engine files, and NVIDIA GPU is leveraged to achieve high-speed inference;

(3) Postprocessing: State prompts (danger/safe) are labeled at the top-left corner of the video stream, and the RGB format is converted to I420 for output;

(4) System control: The communication serial port of the camera host is used to maintain communication with the plasma host or power host. When a danger state is detected, the plasma host or power host will be triggered to pause operation to protect the endoscope from damage.

3. Testing and Clinical Application Validation

3.1 Test Datasets and Evaluation Metrics

To comprehensively verify the engineering performance of the system, the model performance and corresponding metrics were tested based on the constructed two datasets, including:

(1) Detection metrics: Precision, recall, F1-score, average precision (AP);

(2) Classification metrics: Accuracy, precision, recall, F1-score;

(3) Real-time metrics: Processing frame rate, total processing time per frame, time consumption distribution of each module;

(4) Engineering stability metrics: Continuous operation time, anomaly rate, resource occupancy rate.

3.2 Model Test Results

3.2.1 Knife Tip Detection Performance

The detection performance of the YOLOv8 model on the dataset is shown in Figure 6.

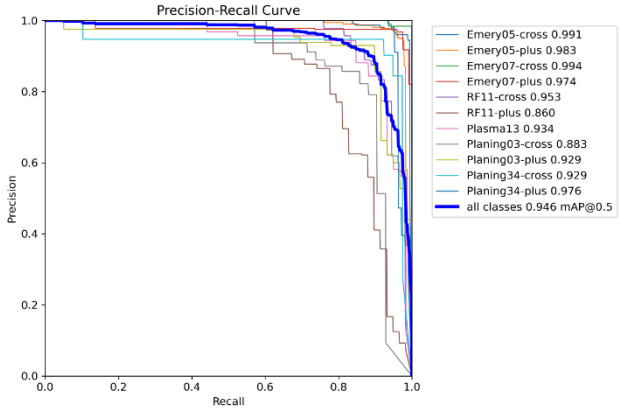

Figure 6 PR curve

The model achieves high-precision detection of knife tips for different types of surgical instruments, with a mAP@50 of 94.6% and a mAP@50:95 of 71.22%. The overall precision is 91.2%, the overall recall is 91.39%, and the overall F1-score is 91.17%, meeting the detection accuracy requirements for engineering applications. The detection data for each type of knife tip are shown in Table 3.

Table 3 Knife Tip Detection Performance (Clinical Dataset)

Instrument Type	Precision (%)	Recall (%)	F1-Score (%)	AP@50 (%)
Emery05-cross	94.46	99.10	96.72	99.14
Emery05-plus	95.49	96.17	95.83	98.34
Emery07-cross	91.07	100.00	95.33	99.41
Emery07-plus	95.90	97.58	96.73	97.45
RF11-cross	85.98	90.74	88.30	95.29
RF11-plus	88.22	70.69	78.49	86.01
Plasma13	88.02	87.15	87.58	93.38
Planing03-cross	85.51	85.71	85.61	88.33
Planing03-plus	90.38	89.83	90.11	92.92
Planing34-cross	90.15	93.93	92.00	92.85
Planing34-plus	98.40	94.36	96.16	97.62

After TensorRT optimization, the inference time per frame of the model is reduced from 22 ms to 6 ms, a 3.7-fold improvement compared with the original model, which guarantees the real-time performance of the system.

3.2.2 Danger Classification Performance

The classification performance of the MobileNetV3-Large model on the dataset is shown in Figure 7.

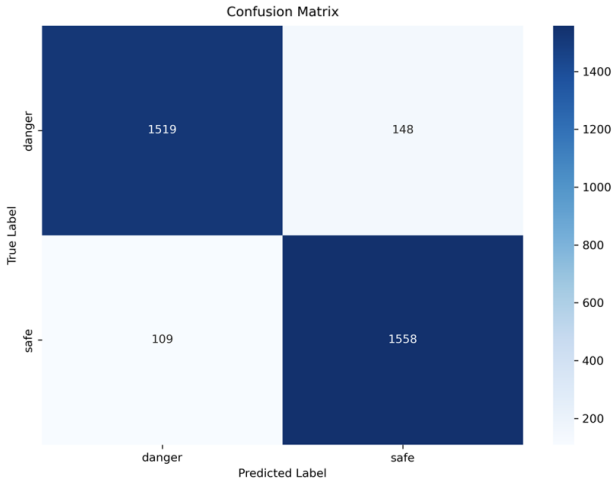

Figure 7 Confusion Matrix

As shown in Table 4, the overall precision of the system reaches 92.31%, and the average recall is 92.29%, ensuring that collision risks are not missed; the overall accuracy is 92.29%, reducing the interference of false alarms on surgeons.

Table 4 Danger State Classification Performance (Clinical Dataset)

Danger Level	Precision (%)	Recall (%)	F1-Score (%)
Dangerous (<3mm)	93.30	91.12	92.20
Safe (≥3mm)	91.32	93.46	92.38
Overall (Macro Average)	92.31	92.29	92.29

On the simulation dataset containing complex scenarios such as bone chip occlusion and bleeding, the system's danger recognition accuracy remains above 90%, indicating its strong engineering robustness.

3.2.3 Real-Time Performance and Resource Occupancy

The real-time performance of the system in the 4K@60fps scenario is shown in Table 5. The overall processing frame

rate of the system reaches 60 FPS, with a total processing time per frame of approximately 16.3 ms. The time consumption distribution of each module is reasonable, meeting the real-time requirements for engineering applications.

Table 5 System Real-Time Performance Test Results

Module	Time per Frame (ms)	Proportion (%)	Engineering Optimization Target
Image Preprocessing	2.5	15.3	< 3 ms
Knife Tip Detection	6.0	36.8	< 7 ms
Danger Classification	2.8	17.2	< 3 ms
Image Post-Processing	2.3	14.1	< 3 ms
System Scheduling	2.7	16.6	< 3 ms
Total	16.3	100	< 17 ms

Resource occupancy tests show that the GPU utilization rate stabilizes at 75%-85% during system operation, the CPU utilization rate is about 40%, and the memory occupancy is about 16 GB, all within the bearing capacity of the hardware platform and without affecting the normal operation of other surgical equipment.

3.2.4 Stability Test

The results of the system continuous operation stability test show no crashes or freezes during 72 hours of continuous operation, with an anomaly rate of 0. In 100 clinical simulated surgeries, the average system operation time is 120 minutes, without data transmission interruption or early warning delay, meeting the stability requirements for engineering applications.

3.3 Clinical Application Verification

Clinical application verification of the system was carried out in 20 minimally invasive spinal surgeries at a tertiary hospital, with the system integrated into the existing surgical endoscopic system. The application effects and surgeons' feedback were recorded as follows:

1. Device compatibility: The system was successfully connected with 4K endoscopes of three mainstream brands, and all surgeries were completed successfully without anti-collision protection system failure caused by endoscope replacement;

2. Anti-collision effectiveness: A total of 163 dangerous states were triggered during the surgeries, and the system responded in a timely manner and realized linkage control of the corresponding equipment, successfully avoiding potential collision risks with no endoscopic damage;

3. Surgeon feedback: 10 orthopedic surgeons participating in the verification gave high evaluations on the ease of use and reliability of the system; 90% of the surgeons considered the early warning information accurate and non-distracting, and 85% expressed their willingness to continue using the system in subsequent surgeries;

4. Cost-effectiveness: Compared with historical data, the collision damage rate of surgical endoscopes was zero during the application period, and it is estimated that the annual equipment maintenance cost can be reduced by about 800 USD per rigid endoscope, with an investment return rate of 300%.

4. Discussion and Optimization Directions

4.1 Advantages

Compared with existing technologies, the engineering advantages of this system are as follows:

1. High real-time performance: Through TensorRT optimization and CUDA parallel computing, the system achieves a processing frame rate of 60 FPS, meeting the real-time processing requirements of 4K endoscopic videos;

2. High reliability: The adoption of a detection-classification two-stage architecture and an existence verification mechanism results in detection accuracy and classification accuracy both exceeding 90%, ensuring the reliability of anti-collision protection;

3. Strong engineering practicality: The visual early warning method conforms to surgeons' operating habits with no invasive impact, and significantly reduces equipment maintenance costs, achieving good economic benefits.

4.2 Limitations and Optimization Directions

Despite the good clinical application effects of the system, there are still several areas for improvement:

1. Model adaptability: The current system supports 11 types of commonly used surgical instruments; in the future, the dataset will be expanded through transfer learning to improve the adaptability to new types of instruments;

2. Distance measurement accuracy: The current distance measurement is based on monocular camera geometric solution; binocular endoscopic data will be introduced in the follow-up to further improve the close-range measurement accuracy;

3. Adaptive threshold: The current danger threshold is fixed; in the future, it will be combined with surgeons' operating habits and surgical types to realize adaptive threshold adjustment;

4. Multi-center verification: Larger-scale, multi-center clinical engineering verification needs to be carried out to further verify the effectiveness of the system in different hospitals and equipment environments;

5. Response speed of serial port linkage equipment: The response speed of plasma and power hosts connected via serial port needs to be optimized based on clinical feedback.

5. Conclusion

Aiming at the clinical problem of endoscopic collision damage in minimally invasive spinal surgery, this paper designs and implements an anti-collision protection system based on YOLOv8 and MobileNetV3. Through a GPU-accelerated image preprocessing pipeline, TensorRT-optimized detection and classification models, and modular engineering design, the system realizes the functions of accurate knife tip detection, danger state recognition and real-time anti-collision protection. Engineering testing and clinical application verification show that the system achieves a knife tip detection mAP@50 of 94.6% and a classification accuracy of 92.29%, with a stable processing frame rate of 60 FPS, fully meeting the application requirements of 4K@60fps endoscopic systems. Clinical application results demonstrate that the system can effectively avoid collision damage between knife tips and endoscopes, significantly reduce the maintenance cost of medical equipment, and has important clinical practical value and broad promotion prospects.

Conflict of Interest

None.

Acknowledgements

None.

References

1. Haihang Fei. Maintenance cost control of rigid endoscopes[J]. Medical Equipment, 2016(02):168-169.
2. Dewei Lu, Lifang Liu. Fault analysis of medical endoscopes and improvement of maintenance management quality[J]. China Medical Equipment, 2019,16(7):187-188.

3. Na Yang, Shi Geng, Tianlei Zheng, et al. Quality control and management measures for rigid endoscopes based on maintenance data analysis[J]. Medical Equipment, 2024,37(21):58-61.
4. Nema S, Vachhani L. Surgical instrument detection and tracking technologies: Automating dataset labeling for surgical skill assessment[J]. Journal of Medical Systems, 2021,45(12):1-14.
5. Jin A, Yeung S, Jopling J, et al. Tool detection and operative skill assessment in surgical videos using region-based convolutional neural networks[C]//IEEE Winter Conference on Applications of Computer Vision. Lake Tahoe, NV, USA: IEEE, 2018:691-699.
6. Choi B, Jo K, Choi S, et al. Surgical-tools detection based on convolutional neural network in laparoscopic robot-assisted surgery[C]//IEEE Engineering in Medicine and Biology Society Annual Conference. Jeju, South Korea: IEEE, 2017:1756-1759.
7. Hasan M K, Calvet L, Rabbani N, et al. Detection, segmentation, and 3D pose estimation of surgical tools using convolutional neural networks and algebraic geometry[J]. Medical Image Analysis, 2021,70:101994.
8. Namazi B, Sankaranarayanan G, Devarajan V. LapTool-Net: A contextual detector of surgical tools in laparoscopic videos using deep learning[J]. Surgical Endoscopy, 2022,36(1):679-688.
9. NVIDIA Corporation. TensorRT Documentation[EB/OL]. https://docs.nvidia.com/deeplearning/tensorrt/overview/index.html, 2023.