Mathematical Biosciences and Engineering

MBE, 21(2): 1844–1856.
DOI: 10.3934/mbe.2024080
Received: 03 October 2023
Revised: 19 December 2023
Accepted: 25 December 2023
Published: 03 January 2024

http://www.aimspress.com/journal/MBE

Research article

Research on workflow recognition for liver rupture repair surgery

Yutao Men[1,3], Zixian Zhao[1,2,3], Wei Chen[1,3], Hang Wu[2], Guang Zhang[2], Feng Luo[2] and Ming Yu[2,*]

[1] Tianjin Key Laboratory for Advanced Mechatronic System Design and Intelligent Control, School of Mechanical Engineering, Tianjin University of Technology, Tianjin 300384, China
[2] Medical Support Technology Research Department, Systems Engineering Institute, Academy of Military Sciences, People's Liberation Army, Tianjin 300161, China
[3] National Demonstration Center for Experimental Mechanical and Electrical Engineering Education, Tianjin University of Technology, Tianjin 300384, China

* **Correspondence:** Email: yuming_1990@outlook.com; Tel: +8615510923627.

Abstract: Liver rupture repair surgery serves as one tool to treat liver rupture, especially beneficial for cases of mild liver rupture hemorrhage. Liver rupture can catalyze critical conditions such as hemorrhage and shock. Surgical workflow recognition in liver rupture repair surgery videos presents a significant task aimed at reducing surgical mistakes and enhancing the quality of surgeries conducted by surgeons. A liver rupture repair simulation surgical dataset is proposed in this paper which consists of 45 videos collaboratively completed by nine surgeons. Furthermore, an end-to-end SA-RLNet, a self attention-based recurrent convolutional neural network, is introduced in this paper. The self-attention mechanism is used to automatically identify the importance of input features in various instances and associate the relationships between input features. The accuracy of the surgical phase classification of the SA-RLNet approach is 90.6%. The present study demonstrates that the SA-RLNet approach shows strong generalization capabilities on the dataset. SA-RLNet has proved to be advantageous in capturing subtle variations between surgical phases. The application of surgical workflow recognition has promising feasibility in liver rupture repair surgery.

Keywords: surgical workflow recognition; liver rupture repair surgery; attention mechanism; recurrent convolutional network; deep learning; image classification

1. Introduction

It has been demonstrated for various minimally invasive gastrointestinal surgeries that there is a close association between intraoperative technical presentation as assessed by video studies and patient complications [1]. Recognizing workflow stages from endoscopic surgical videos is essential for acquiring indicators that deliver the quality, efficacy and consequence of the surgery and providing insights into surgical group skills [2]. Accurate recognition of the surgical workflow is indispensable for constructing a context-aware system for computer-assisted intervention (CAI) [3]. To advance the therapeutic effect on patients, modern operating rooms require context-aware systems that monitor the surgical procedure and enhance surgeon interaction [4,5]. As a fundamental component in building a sophisticated assistive system for the operating room [6], the identification of surgical process can help the system in real-time monitoring and optimizing the surgical workflow. This skill can help surgeons make decisions, reduce surgical errors and warn them of potential complications [7,8].

Regarding surgical workflow identification, several deep learning (DL) methods are proposed in the paper, including spatial and temporal models. Temporal information along the surgical video sequence is crucial for surgical workflow recognition [9]. In 2017, Twinanda et al. presented a convolutional neural network (CNN) model called Endonet, which was designed for the dual tasks of surgical workflow recognition and tool classification [10]. To overcome the shortcomings of statistical models, subsequent approaches have utilized long short-term memory (LSTM) networks to learn temporal features [11,12]. This approach aimed to enhance the capturing of temporal correlations throughout surgical workflows by utilizing LSTM. Jin et al. introduced SV-RCNet, a DL framework that integrates a CNN and recurrent neural network (RNN) in a distinctive recursive convolutional architecture. This framework, named SV-RCNet, included end-to-end training and prior knowledge reasoning and was proposed for workflow identification in surgical videos. The goal of SV-RCNet was to effectively utilize the complementary information obtained from visual and temporal features gathered from surgical videos [11]. Similarly, Jin et al. introduced the MTRCNet-CL approach, which successfully took advantage of the link between surgical tool detection and surgical workflow recognition. Performance on both tasks was enhanced by using this approach. In addition, Jin et al. introduced an innovative loss function to take into account the relationship between workflows and tools [12]. Jalal et al. proposed a DL model for surgical workflow identification and prediction, which combined CNN and a nonlinear autoregressive network with exogenous inputs [13]. Transformers have been successfully applied in clinical tasks such as image synthesis, reconstruction, segmentation, detection and diagnosis [14]. For example, OperA, a transformer-based model that accurately predicted surgical phases from long video sequences, was introduced by Czempiel et al. [15]. For the first time in surgical process analysis, Gao et al. suggested using a transformer to reconsider the complementary effect of spatial and temporal features for precise surgical phase recognition [16].

In addition, deep learning helps doctors diagnose oral cancer through image recognition technology [17]. Deep learning helps doctors in the diagnosis of brain tumors [18]. Despite the fact that various DL methods have been extensively used in surgical process analysis [19], these techniques have not yet been applied to liver rupture repair surgery. The liver is prone to injury due to its size, weight and delicate texture. It also has a rather complex structure and function and contains a large number of vital blood vessels. Due to this, liver rupture can result in several

consequences, such as hemorrhagic shock, biliary peritonitis and secondary infections. The safety of patients' lives is seriously threatened by these problems. Currently, the main treatments for liver injuries are suturing and wound debridement. However, due to the difficulty in obtaining real liver rupture repair surgery videos, a simulated liver rupture repair surgery dataset was created by our research group. The dataset is referred to as "liver45" in this paper. It comprises a total of 45 simulated surgery videos. Our main contributions are summarized as follows:

1) The liver rupture repair simulation surgery utilizes an organ model that closely resembles human tissue for surgical operation. This simulated surgery data set was created through the joint efforts of nine surgeons, with each carrying out five simulated surgeries, yielding a total of 45 simulated surgical videos as data.

2) The network, SA-RLNet, is specifically engineered to outshine in the area of surgical workflow recognition. Unlike ResNet50-LSTM, SA-RLNet effectively leverages time series information. This network adopts an end-to-end structure, using ResNet50 to extract visual features. The inclusion of an LSTM network enables the learning of temporal features, while a self-attention mechanism automatically identifies the significance of input features at different instances.

3) By using a dataset that our research group has created, we have effectively proved the viability of SA-RLNet in recognizing surgical processes during simulated liver rupture repair procedures. This validation demonstrates how SA-RLNet may be used in practical surgical circumstances.

2. Materials and methods

2.1. Data acquisition

The simulated surgery for liver rupture repair dataset was obtained from the damage control surgery database established by the research group. It utilizes simulated human models, created by the project team, as the subjects of the trial. The model consists of simulated organs such as the liver, intestine, kidney and abdominal simulated skin, which are all made from biological materials like hydro-gels, with a high resemblance to human tissue. After institutional testing, key parameters of the simulated human organs, such as water content, elastic modulus and electrical conductivity, have a deviation of no more than 20% compared to actual human organs. These similarities allow them to duplicate touch sensations and provide force feedback similar to those of the human body. Three surgical procedures, namely, liver rupture repair, intestinal anastomosis and temporary abdominal closure, were performed on these simulated human models by a team of nine surgeons for experimentations included in the database.

The project team captured three types of surgical procedures by surgical coaxial cameras to get the surgical videos with RGB color images. The database of simulated liver rupture repair surgeries was generated collaboratively by nine surgeons. Each surgeon performed five simulated surgeries, resulting in a total of 45 instances of simulated surgical video data. These videos were recorded using the surgical coaxial camera at an angle that nearly matched the angle of view of the surgeons. The primary objective of the self-developed simulated liver rupture repair surgery is to manage simulated liver rupture and hemorrhage. The workflow of the simulated surgery is consistent with actual liver rupture repair operations, despite the simulated liver's internal structure being somewhat simplified and lacking vascular systems. Five types of tools are needed for this mock surgery, including clamps,

forceps, scissors, holders and sponges. Sponge is mainly used to fill ruptures and control bleeding. The five types of tools are illustrated in Figure 1.

Figure 1. Five types of surgical tools in liver rupture repair surgery.

Under the guidance of surgical experts, the surgical workflow for liver rupture repair has been divided into four distinct phases based on the operations and tools used in each phase. The details of these phases are described comprehensively in Table 1. The scene examples of each surgical phase of the simulated liver rupture repair surgery are shown in Figure 2. The workflow sequence of the simulated liver rupture repair surgery is shown in Figure 2.

Table 1. Surgical Phase Division. Transverse mattress suture: A suturing technique of mattress suture parallel to the incision. Interrupted suture: The needle passes through the skin, epidermis and dermis and traverses the subcutaneous tissue across the incision to the opposite side of the skin, with each stitch tied individually.

ID	Phase	Annotations
P1	Cleaning and hemostasis	The forceps and clamp are used to clean necrotic tissue and blood clots and fill the ruptured site with a gelatin sponge to control hemorrhage.
P2	Suture-P1	The method of transverse mattress suture is used to preliminarily manage the wound.
P3	Suture-P2	The interrupted suture is utilized on the outer side of the transverse mattress sutures to smooth the wound.
P4	Checking	Check the sutures to assess if any supplementary stitches are required.

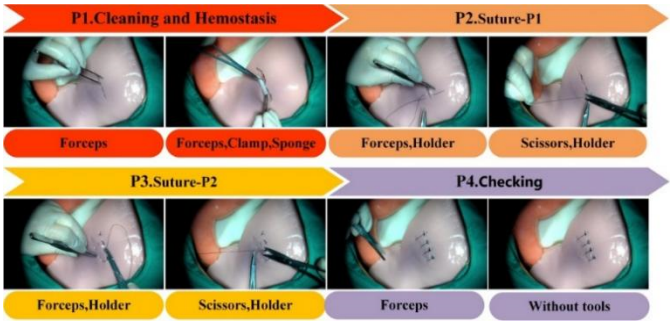

Figure 2. Examples of surgical phase scenarios and the tools employed.

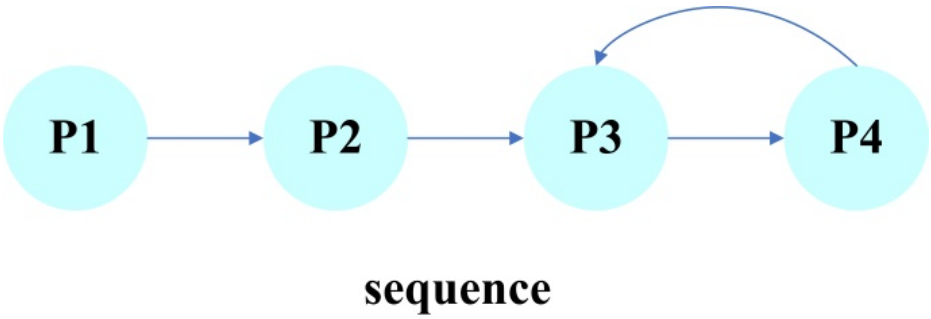

Figure 3. The sequence of surgical workflow. The meaning of the curved arrow: After completing the P4 checking, it is found that the suturing condition at the liver rupture is unsatisfactory, which requires a refill needle step (P3).

2.2. Data preprocessing

Manual annotations are made for every frame of the simulated surgical video of liver rupture repair. For testing and evaluation purposes, the liver45 videos are divided at random. The specific division is as follows: A training dataset contains 25 videos, a validation dataset contains 12 videos, and a testing dataset contains eight videos. The initial frame rate of the surgical videos is 30 frames per second (30 fps). To reduce the computational burden, down-sampling processing is employed to decrease the frame rate to one frame per second.

As can be seen in Table 2, the duration of the surgical phase of Checking (P4) is relatively short, leading to fewer samples in the training dataset. This issue is solved by adjusting the frame rate, from 30 fps to 3 fps, in P4. After down-sampling, the training dataset consists of 12,460 images from 25 surgical videos, the validation dataset consists of 6146 frames of images from 12 surgical videos, and the testing dataset consists of 3718 frames from eight surgical videos. The original resolution of video frames is adjusted from 1920 × 1080 to 224 × 224, and data preprocessing is done. To accelerate the convergence rate of the model, the mean and standard deviation of the training dataset images are calculated, and the results are mean (0.3436, 0.3807, 0.3728) and standard deviation (0.3418, 0.3384, 0.3358). The PyTorch API is used to perform data augmentation to avoid over-fitting. This approach provides real-time data augmentation and reduced memory consumption. Surgical videos are a kind of sequential data, so the continuous video frames obtained after down-sampling processing are also sequential data. This paper adopts the data processing method proposed by Pan et al. [20]. As illustrated in Figure 4, a sliding window of size five is used to generate sequential data, moving backward one frame at a time. There is a 4-frame overlap between adjacent sequences, with the last frame being updated. During network training, it is necessary to randomly shuffle these sequences to avoid the gradient being too extreme when the network weights are updated, which can potentially lead to over-fitting of the final model [20].

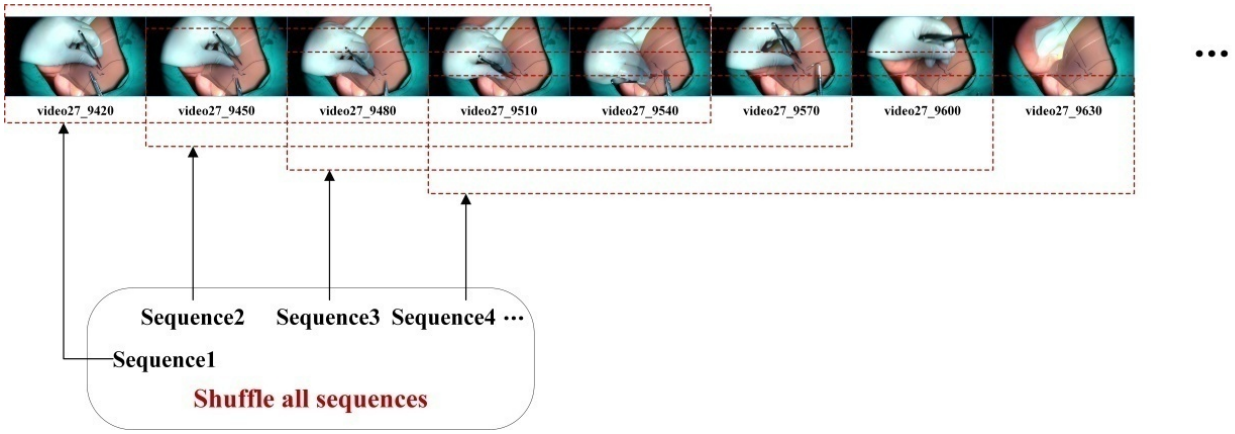

Figure 4. Data processing method. Please note that the "27" within "video27_9420" represents the number of the surgical video, and "9420" represents frame 9420 of the surgical video.

Table 2. Phases and time statistics of 45 surgical videos.

ID	Phase	Duration (seconds)
P1	Cleaning and hemostasis	67.6 ± 31.72
P2	Suture-P1	248.33 ± 75.94
P3	Suture-P2	161.58 ± 53.05
P4	Checking	6.27 ± 4.09

2.3. Training of the networks

This paper conducts training on six distinct networks: EfficientNetV2-S [21], ShuffleNetV2-1× [22], MobileNetV3-large [23], ResNet50, ResNet50-LSTM and SA-RLNet. Long short-term memory (LSTM) is a special type of recurrent neural network (RNN) that is excellent in handling long-term dependence problems [24]. ResNet50-LSTM is an end-to-end structure that integrates ResNet50 with LSTM. This combination makes it highly suitable for effectively processing sequential data.

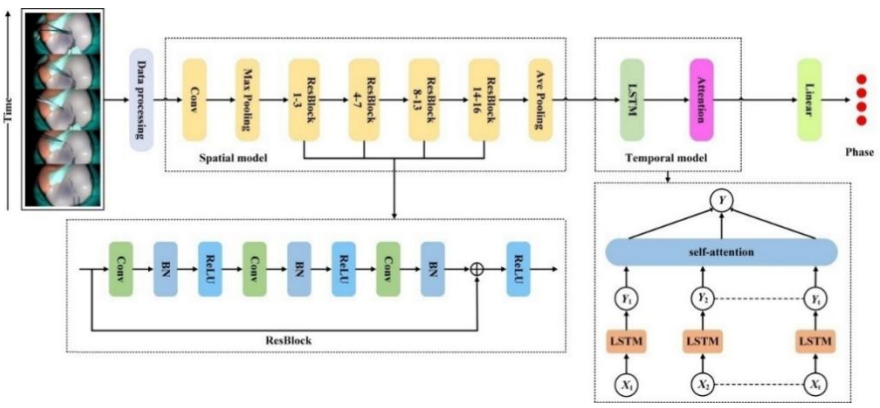

Figure 5. SA-RLNet.

As for the SA-RLNet (Figure 5), Jin et al. verified that ResNet50 performs well in feature extraction [11]. Therefore, ResNet50 is chosen as the appropriate feature extractor. In SA-RLNet, memory units in the LSTM are used to capture temporal information of past frames. To fairly distribute the attention weights of each input feature, the self-attention mechanism is proposed, which can automatically distinguish the importance of input features at various moments and establish the links between input features. The SA-RLNet consists of 49 convolutional layers, an LSTM layer (including 5 LSTM units and 128 hidden units), an attention layer and a fully connected layer (acting as the output layer). The self-attention mechanism is employed in the attention layer of the SA-RLNet [15]. The output of the LSTM layer is used as the input of the self-attention layer, and the query Q, key K and value V are obtained through three linear layers. Then, the vector weighted by the attention mechanism is determined according to Eq (1) (dk = 128). Finally, the prediction results are output through a fully connected layer acting as the output layer.

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{1}$$

The experimental environment is based on Python 3.8. The framework of this study is based on PyTorch 1.8.0 and trained on a device equipped with a 24 GB TITAN RTX GPU. To demonstrate the improvement of the self-attention module on the performance of the ResNet50-LSTM network in surgical procedure recognition, this study maintains consistent loss function, optimizer and hyper-parameter Settings between the ResNet50-LSTM and the SA-RLNet. Specifically, the cross entropy loss function is used as the loss function. Stochastic gradient descent (SGD) is used as the optimizer. The value of the initial learning rate for the convolutional layer is 0.0001, for the LSTM is 0.001 and for the output layer of the linear layer is 0.001. The number of attenuation-learning rate rounds of the optimizer is 15, the attenuation-multiplier of the learning rate is 0.1, the other parameters of the optimizer keep their default values, and momentum is 0.9. The batch size of the training images is 13, and the total number of training rounds is 25. Taking advantage of the generalization ability of transfer learning, the ResNet50 module of the three networks trained is initialized by the weights trained on the ImageNet dataset.

2.4. Evaluation metrics

To quantitatively analyze the performances of the six trained networks in this paper, four different indexes are adopted: precision (Eq (2)), recall (Eq (3)), accuracy and F1 score (Eq (4)). These indexes are the most commonly used standards for evaluating the effectiveness of workflow recognition models. The ground truth set of the phase is represented by GT, and the prediction set of the phase is represented by P. First, precision, recall and F1 score values of each phase are calculated in this paper. Subsequently, these indexes for all phases are averaged to obtain corresponding values for the entire test dataset. Accuracy is assessed on the entire test dataset, which is defined as the percentage of correct detection in the entire test dataset. However, accuracy cannot accurately evaluate the generalization ability of the model if the amount of data in the classified samples is not balanced. Therefore, the F1 score index is introduced in this paper. F1 score can be considered as a weighted average of model precision and recall which can objectively evaluate the recognition ability of the model for each surgical phase.

$$Precision = \frac{|GT \cap P|}{|P|} \tag{2}$$

$$Recall = \frac{|GT \cap P|}{|GT|} \tag{3}$$

$$F1\ score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{4}$$

3. Results

3.1. Phase detection

Table 3. Evaluation results of the SA-RLNet method in four phases (the "sample" column represents the sample size for that category).

ID	Phase	Precision (%)	Recall (%)	F1-score (%)	Sample
P1	Cleaning and hemostasis	86.1	95.4	90.5	544
P2	Suture-P1	95.8	91.9	93.8	1751
P3	Suture-P2	87.9	92.8	90.3	1258
P4	Checking	68.5	44.8	54.2	175
	Average	84.6	81.2	82.9	

As can be seen from Table 3, in the three phases of P1–P3, the network shows a good recognition effect, but a poor one in the checking (P4).

3.2. Baseline comparison

The mean values of precision, recall and F1-score of the six networks at each phase for the test dataset are presented in Table 4. Four videos are randomly selected from the test dataset for the visual recognition effect, as shown in Figure 6. ResNet50-LSTM outperforms ResNet50 by 3.8% and 0.9% in Accuracy and F1-score. This shows the importance of temporal information for the recognition of surgical phases. SA-RLNet still outperforms ResNet50-LSTM by 2.2% and 0.4% in accuracy and F1-score.

Table 4. Phase recognition results.

Method	Precision (%)	Recall (%)	Accuracy (%)	F1-score (%)
ShuffleNetV2-1× [22]	81.8	81.2	86.1	81.5
MobileNetV3-large [23]	78.1	75.3	83.3	76.7
EfficientNetV2-S [21]	83.3	71.9	83.6	77.2
ResNet50	81.5	81.8	84.6	81.6
ResNet50-LSTM	85.3	79.8	88.4	82.5
SA-RLNet	84.6	81.2	90.6	82.9

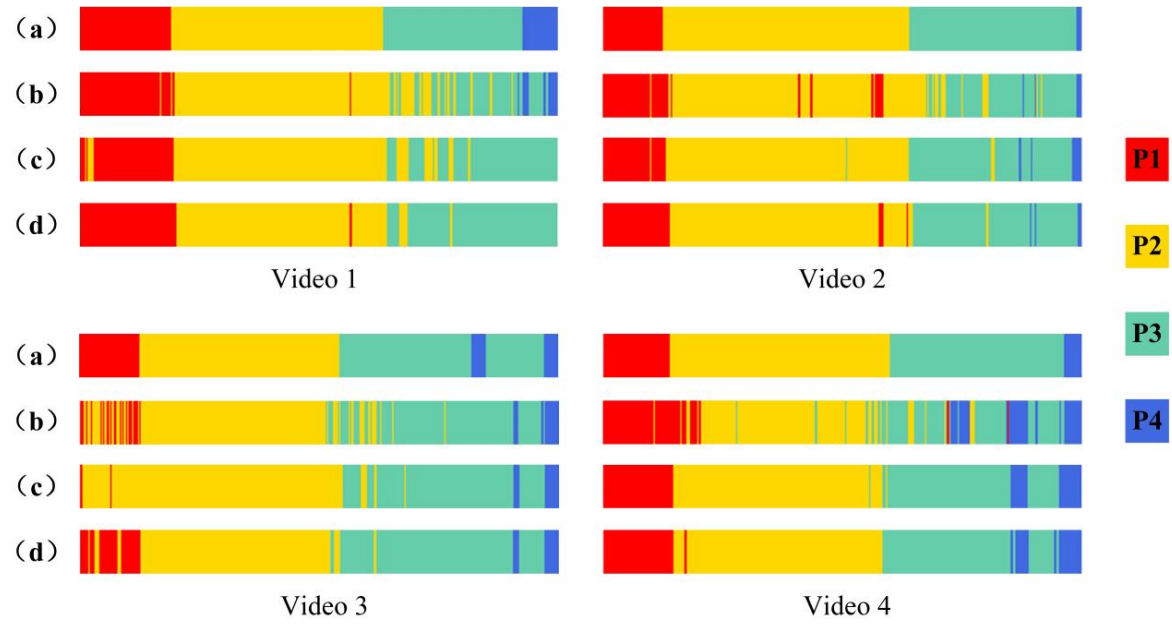

Figure 6. Qualitative results of phase recognition effect. (a) Ground truth, (b) ResNet50, (c) ResNet50-LSTM, (d) SA-RL Net. Phase labels are indicated by P1 to P4.

In Figure 7, the horizontal axis represents the number of video frames, and the vertical axis represents the phase labels (P1–P4). The ground truth of the phase is shown in red, and the prediction results are shown in blue. More overlapping parts in the prediction results indicate higher prediction accuracy. The sequence containing the transition frame shows the phase in which the current image is located and the surgical tools included in the image.

A sequence containing the transition video frames from P1 to P2 is shown in Figure 7. In this sequence, four video frames in P1 contain three surgical tools, which are forceps, sponge and clamp, while one video frame in P2 contains only the sponge tool. During the transition from P1 to P2, the surgical tools that exist in the video frame have changed, so the characteristics of the input will change accordingly. To accurately recognize the transition frames, it is necessary to introduce the self-attention mechanism to assign weights to each input feature in the video frame sequence. Similarly, the surgical tools that exist in the video frame change during the transition from P2 to P3, as well as from P3 to P4. This means that during these transitions, input features change, and it is also necessary to introduce the self-attention mechanism to recognize the transition frames accurately.

Ablation experiments have been used to train and assess multiple innovative networks, such as EffectiveNetV2_S [21]. The outcomes validate SA-RLNet's exceptional ability to precisely identify surgical procedures. In conclusion, this research has a lot of potential to enhance surgical practices by improving workflow recognition, which would eventually improve patient outcomes.

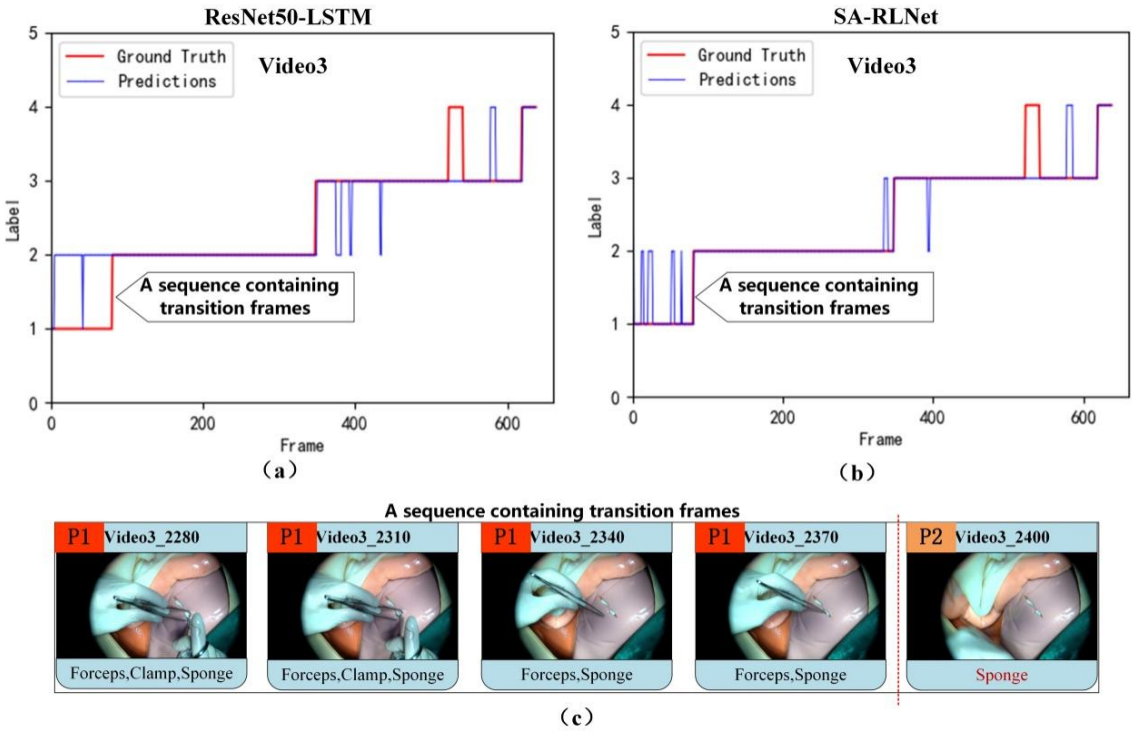

Figure 7. Prediction results of a complete surgical video through (a) ResNet50-LSTM and (b) SA-RLNet. (c) A sequence containing transition frames.

4. Discussion

This study introduces a dataset called liver 45, which consists of 45 surgical videos performed by nine surgeons. A self-attention mechanism-based approach, called SA-RLNet, is proposed in this paper to apply surgical workflow recognition in liver rupture repair surgery. ResNet50 is used as the feature extractor in the SA-RLNet method, and the memory units in the LSTM are used to memorize the temporal information of past frames. In this paper, the self-attention mechanism is introduced to automatically recognize the importance of input features at different moments and associate the relationships between input features. The results show that the SA-RLNet approach has a better effect on surgical workflow recognition on the liver 45 dataset. This proves that our method can effectively capture the characteristics of different phases in liver rupture repair surgery and accurately recognize them.

Although the improvement of the SA-RLNet in accuracy and F1-score is not significant compared with the other networks, the network performs better in the recognition of the transitional video frames between the two surgical phases. This observation suggests that the SA-RLNet network has advantages in capturing the subtle changes between phases during surgery. The SA-RLNet may be more sensitive to subtle variations in the surgical workflow of surgery and can more accurately recognize the specific phases of the transitional frames. Despite the limited improvement in overall performance, the advantages of the SA-RLNet in the recognition of transitional video frames may still be of great significance for some specific application scenarios, especially in situations where accurate recognition of surgical phase transitions is required.

Through the analysis of the dataset, two possible reasons for the poor recognition in Checking (P4) are inferred. First, the sample images of P4 have low similarity, which may increase the difficulty of recognition. Second, the sample size in the Checking (P4) phase is still low despite having increased the frame extraction rate in the Checking (P4) phase and using data augmentation during model loading. This insufficient sample size limits the learning ability of the network in P4. As for the sequence of the surgical workflow, two approaches are illustrated in Figure 3. For example, after completing the P4 Checking, it is found that the suturing condition at the liver rupture is unsatisfactory, which requires a refill needle step (P3). Consequently, P3 can occur after P2 or P4, which adds to the difficulty of surgical workflow recognition.

5. Conclusions

In conclusion, this study demonstrates the feasibility of applying deep learning technology for surgical workflow recognition in liver rupture repair surgery. The simulated surgical scenarios are slightly different from real surgical scenarios due to the simple internal structure of the simulated liver and the absence of vascular structures, and therefore the simulated surgical scenarios have no blood contamination. This is a limitation of our dataset and a realistic problem that needs to be addressed in subsequent study.

Use of AI tools declaration

The authors declare that no artificial intelligence tools were used in the writing of the paper.

Conflict of interest

The authors declare there is no conflict of interest.

References

1. L. S. Feldman, A. D. Pryor, A. Gardner, B. Dunkin, L. Schultz, M. Awad, et al., Sages video-based assessment (vba) program: A vision for life-long learning for surgeons, *Surg. Endoscopy*, **34** (2020), 3285–3288. https://doi.org/10.1007/s00464-020-07628-y
2. B. Zhang, J. Abbing, A. Ghanem, D. Fer, J. Barker, R. Abukhalil, et al., Towards accurate surgical workflow recognition with convolutional networks and transformers, *Comput. Methods Biomech. Biomed. Eng.: Imaging Visualization*, **10** (2022), 349–356. https://doi.org/10.1080/21681163.2021.2002191
3. O. Dergachyova, D. Bouget, A. Huaulmé, X. Morandi, P. Jannin, Automatic data-driven real-time segmentation and recognition of surgical workflow, *Int. J. Comput. Assisted Radiol. Surg.*, **11** (2016), 1081–1089. https://doi.org/10.1007/s11548-016-1371-x
4. L. Maier-Hein, S. Vedula, S. Speidel, N. Navab, R. Kikinis, A. Park, et al., Surgical data science for next-generation interventions, *Nat. Biomed. Eng.*, **1** (2017), 691–696. https://doi.org/10.1038/s41551-017-0132-7
5. N. Bricon-Souf, C. R. Newman, Context awareness in health care: A review, *Int. J. Med. Inf.*, **76** (2007), 2–12. https://doi.org/10.1016/j.ijmedinf.2006.01.003

6. N. Padoy, Machine and deep learning for workflow recognition during surgery, *Minimally Invasive Ther. Allied Technol.*, **28** (2019), 82–90. https://doi.org/10.1080/13645706.2019.1584116
7. A. Huaulmé, P. Jannin, F. Reche, J. Faucheron, A. Moreau-Gaudry, S. Voros, Offline identification of surgical deviations in laparoscopic rectopexy, *Artif. Intell. Med.*, **104** (2020). https://doi.org/10.1016/j.artmed.2020.101837
8. B. Zhang, A. Ghanem, A. Simes, H. Choi, A. Yoo, Surgical workflow recognition with 3DCNN for Sleeve Gastrectomy, *Int. J. Comput. Assisted Radiol. Surg.*, **16** (2021), 2029–2036. https://doi.org/10.1007/s11548-021-02473-3
9. C. Garrow, K. Kowalewski, L. Li, M. Wagner, M. Schmidt, S. Engelhardt, et al., Machine learning for surgical phase recognition: A systematic review, *Ann. Surg.*, **273** (2021), 684–693. https://doi.org/10.1097/SLA.0000000000004425
10. A. Twinanda, S. Shehata, D. Mutter, J. Marescaux, M. de Mathelin, N. Padoy, EndoNet: A deep architecture for recognition tasks on laparoscopic videos, *IEEE Trans. Med. Imaging*, **36** (2017), 86–97. https://doi.org/10.1109/TMI.2016.2593957
11. Y. Jin, Q. Dou, H. Chen, L. Yu, J. Qin, C. Fu, et al., SV-RCNet: Workflow recognition from surgical videos using recurrent convolutional network, *IEEE Trans. Med. Imaging*, **37** (2018), 1114–1126. https://doi.org/10.1109/TMI.2017.2787657
12. Y. Jin, H. Li, Q. Dou, H. Chen, J. Qin, C. Fu, et al., Multi-task recurrent convolutional network with correlation loss for surgical video analysis, *Med. Image Anal.*, **59** (2020). https://doi.org/10.1016/j.media.2019.101572
13. N. Jalal, T. Alshirbaji, K. Möller, Predicting surgical phases using CNN-NARX neural network, *Curr. Dir. Biomed. Eng.*, **5** (2019), 405–407. https://doi.org/10.1515/cdbme-2019-0102
14. K. He, C. Gan, Z. Li, I. Rekik, Z. Yin, W. Ji, et al., Transformers in medical image analysis, *Intell. Med.*, **3** (2023), 59–78. https://doi.org/10.1016/j.imed.2022.07.002
15. T. Czempiel, M. Paschali, D. Ostler, S. Tae Kim, B. Busam, N. Navab, Opera: Attention-regularized transformers for surgical phase recognition, in *Medical Image Computing and Computer-Assisted Intervention*, Springer, (2021), 604–614. https://doi.org/10.1007/978-3-030-87202-1_58
16. X. Gao, Y. Jin, Y. Long, Q. Dou, P. Heng, Trans-SVNet: Accurate phase recognition from surgical videos via hybrid embedding aggregation transformer, *arXiv preprint*, (2021), arXiv:2103.09712. https://doi.org/10.48550/arxiv.2103.09712
17. S. Panigrahi, R. Bhuyan, K. Kumar, J. Nayak, T. Swarnkar, Multistage classification of oral histopathological images using improved residual network, *Math. Biosci. Eng*., **19** (2022), 1909–1925. https://doi.org/10.3934/mbe.2022090
18. A. Hassan, J. Wu, M. Muhammad, U. Muhammad, Brain tumor classification in MRI image using convolutional neural network, *Math. Biosci. Eng.*, **17** (2020), 6203–6216. https://doi.org/10.3934/mbe.2020328
19. D. Birkhoff, A. van Dalen, M. Schijven, A review on the current applicationsof artificial intelligence in the operating room, *Surg. Innovation*, **28** (2021), 611–619. https://doi.org/10.1177/1553350621996961
20. X. Pan, X. Gao, H. Wang, W, Zhang, Y. Mu, X. He, Temporal-based swin transformer network for workflow recognition of surgical video, *Int. J. Comput. Assisted Radiol. Surg.*, **18** (2023), 139–147. https://doi.org/10.1007/s11548-022-02785-y

21. M. Tan, Q. Le, EfficientNetV2: Smaller models and faster training, *arXiv preprint*, (2021), arXiv:2104.00298. https://doi.org/10.48550/arxiv.2104.00298
22. N. Ma, X. Zhang, H. Zheng, J. Sun, ShuffleNet V2: Practical guidelines for efficient CNN architecture design, *arXiv preprint*, (2018), arXiv:1807.11164. https://doi.org/10.48550/arxiv.1807.11164
23. A. Howard, M. Sandler, G. Chu, L. Chen, B. Chen, M. Tan, et al., Searching for MobileNetV3, *arXiv preprint*, (2019), arXiv:1905.02244. https://doi.org/10.48550/arxiv.1905.02244
24. J. Donahue, L. Hendricks, M. Rohrbach, S. Venugopalan, S. Guadarrama, K. Saenko, et al., Long-term recurrent convolutional networks for visual recognition and description, *IEEE Trans. Pattern Anal. Mach. Intell.*, **39** (2017), 677–691. https://doi.org/10.1109/TPAMI.2016.2599174

AIMS Press