中文摘要：

腹腔镜手术场景下的手术阶段划分任务的目标是判断腹腔镜手术视频中每一帧对应的手术阶段。作为计算机辅助医疗系统中最基本的任务之一，手术阶段划分在手术期间以及手术后都有较为广泛的应用。在手术中，手术阶段划分方法可以帮助医生判断当前正在执行的手术阶段，从而让医生对当前手术进度有更加直观的了解，提升手术的效率以及安全性。在手术后，手术阶段划分可用于医疗教育与手术数据的归档。

目前大多数手术阶段划分方法为两阶段架构，即先通过特征提取网络对手术视频进行特征提取，再通过各类时序模型提取时序信息并给出手术阶段划分结果。然而在手术阶段划分方法的优化与改进方面，这些方法大多聚焦于对特征提取网络的提升以及对时序模型的提升，而忽略了手术视频帧中包含的更多可以帮助模型进行手术阶段划分的信息，比如手术视频帧中的器官信息、关键血管信息、特殊手术器械信息等等。相比于手术视频帧中的器官信息，血管信息的可辨认程度更低，器械信息的泛化性更差。因此，本文选择手术视频帧中的器官信息作为附加信息，帮助模型进行手术阶段划分任务。

使用器官信息辅助进行手术阶段划分任务是具有挑战性的。第一个挑战是如何进行手术视频中器官信息的提取。为了对腹腔镜手术视频中器官的位置信息进行提取，需要一个适用于器官检测的腹腔镜手术图像数据集。然而，目前已开放的腹腔镜手术数据集中，缺少符合要求的数据集。此外，手术场景的复杂性，例如光照变化、遮挡、镜头移动等，也使得器官信息的提取变得极其困难。第二个挑战是如何让器官信息辅助手术阶段划分方法。如何将手术视频中得到的器官信息与特征提取网络得到的视频特征融合在一起是实现辅助效果的关键。另外，并不是所有的器官信息都对手术阶段划分有帮助，如何屏蔽掉这些不相关的信息也是具有挑战性的。

针对上述的两个挑战，本文给出了解决方案。首先，针对第一个挑战，本文通过两种不同的方式建立了两个适用于器官检测的腹腔镜手术图像数据集。第一种方式中，本文从腹腔镜手术视频数据集中手动选择合适的视频帧组成数据集，并对每一张视频帧中出现的器官进行手动标注，标注方式为用检测框标记出器官的位置。而在第二种方式中，本文直接基于已公开的器官分割数据集生成器官检测数据集，保留原数据集中的部分视频帧，并将每一张视频帧中对应器官的掩码标注转化为检测框标注。此外，为了解决提出的两个数据集中手术图像数量较少的问题，本文提出了一个基于镜头移动方式的适用于腹腔镜手术图像的数据增强方法。随后，本文在这两个经过数据增强的腹腔镜手术图像数据集上训练了器官检测模型。实验结果表明，器官检测模型可以很好的检测出数据集中器官的位置，并且对图片的亮度以及影响器官的遮挡等均具有很好的鲁棒性。

其次，针对第二个挑战，本文提出了基于器官检测的腹腔镜手术阶段划分方法。此方法由两部分构成，器官检测信息融合方法以及阶段屏蔽的器官信息融合训练方式。其中器官检测信息融合方法旨在将器官检测模型得到的检测结果与特征提取方法得到的手术视频特征相融合。而阶段屏蔽的器官信息融合训练方式则旨在屏蔽融合过程中冗余的器官信息，提高手术阶段划分的性能。随后，本文在腹腔镜手术数据集上进行了一系列的对比实验。实验结果表明，本文所提出的方法可以在多种时序建模架构下适用，且有着符合预期的性能表现。同时，本文在腹腔镜手术数据集上进行了消融实验。实验结果表明，本文所提出方法的两个组成部分均对手术阶段划分有一定帮助。

英文摘要：

The objective of the surgical phase recognition task under the laparoscopic surgery scenario is to determine the surgical phase corresponding to each frame in the laparoscopic surgery video. As one of the most fundamental tasks in computer-assisted medical systems, surgical phase recognition has broad applications both during and after surgery. During the operation, the surgical phase recognition method can assist doctors in determining the current surgical phase being performed, thereby providing a more intuitive understanding of the current progress of the surgery and enhancing the efficiency and safety of the operation. Post-surgery, the surgical phase recognition can be utilized for medical education and archiving of surgical data.

Most current surgical phase recognition methods adopt a two-stage architecture, which first extracts features from the surgical video through a feature extraction network, and then employs various temporal models to extract temporal information and provide surgical phase recognition results. However, in optimizing and improving surgical phase recognition methods, these approaches tend to focus on enhancing the feature extraction network and the temporal models, often overlooking additional information contained in the surgical video frames that could assist the model in surgical phase recognition, such as organ information, key vascular information and special surgical instrument information. Compared to organ information in surgical video frames, vascular information is difficult to recognize, and instrument information has poorer generalization. Therefore, this thesis employs organ information from the surgical video frames as supplementary information to aid the model in the task of surgical phase recognition.

Using organ information to assist in the task of surgical phase recognition is challenging. The first challenge lies in how to extract organ information from surgical videos. To extract the positional information of organs in laparoscopic surgery videos, a laparoscopic surgery image dataset suitable for organ detection is required. However, among the currently open laparoscopic surgery datasets, there is no datasets that meet our requirements. Additionally, the complexity of the surgical scene, such as changes in illumination, occlusion, camera movement, etc., makes the extraction of organ information extremely difficult. The second challenge is how to utilize organ information to aid surgical phase recognition methods. How to integrate the organ information derived from the surgical videos with the video features extracted by the feature extraction network is key to achieving the desired auxiliary effect. Furthermore, not all organ information is helpful for surgical phase recognition. How to filter out this irrelevant information also poses a challenge.

Addressing the mentioned two challenges, this thesis proposes solutions. Firstly, we constructed two laparoscopic surgery image datasets suitable for organ detection in two different ways. For the first dataset, appropriate video frames were manually selected from a laparoscopic surgery video dataset, and the organs appearing in each video frame were manually annotated, with detection boxes marking the location of the organs. For the second dataset, we directly generated an organ detection dataset based on publicly open organ segmentation dataset. We retained some video frames from the original dataset, and converted the mask annotations of the corresponding organs in each video frame into detection box annotations. In addition, to solve the issue of fewer surgical images in the proposed datasets, we introduced a data augmentation method for laparoscopic surgery images based on camera movement. Subsequently, organ detection models were trained on these two enhanced laparoscopic surgery image datasets. Experimental results demonstrate that the organ detection models can effectively detect the location of organs in the videos, exhibiting robustness to variations in brightness and organ occlusion.

Secondly, this thesis proposes a laparoscopic surgery phase recognition method based on organ detection. This method consists of two parts: an organ-detection fusion method and a phase-shielding training approach. The organ-detection fusion method aims to integrate the results obtained from the organ detection model with the video features derived from the feature extraction method. The phase-shielding training approach aims to eliminate redundant organ information in the fusion process, thereby improving the performance of surgical phase recognition. Subsequently, a series of comparative experiments were conducted on the laparoscopic surgery dataset. The experimental results demonstrate that the method proposed is applicable under various temporal modeling architectures and performs as expected. Additionally, ablation experiments were performed on the laparoscopic surgery dataset. The results indicate that both components of the proposed method provide certain aid in surgical phase recognition.