Tool Presence Detection and Feature Extraction in Laparoscopy Videos

Sisira C Sunny[1], Anu Jose[2]

PG Student, Dept. of Computer Science and Engineering,Viswajyothi College of Engineering and Technology, Vazhakulam[1]

Asst. Professor, Dept. of Computer Science and Engineering,Viswajyothi College of Engineering and Technology, Vazhakulam[2]

sisiracs@gmail.com [1], anujose@vjcet.org [2]

Abstract

Medical surgeries are really complex in terms of effort, tools, and procedures. There are different categories of surgeries that differ dramatically in terms of operation. Each category of surgery is different from one another. Tools and Phases or workflows are different. Digital assistance improves the efficiency of such surgeries. And they reduce the cost also. This paper proposes a deep learning approach for surgical tool presence detection and feature extraction for phase recognition. This approach utilizes Convolutional Neural Networks (CNN) , more specifically a fine-tuned VGG16 architecture. Discriminating features are extracted with the help of a CNN. This feature helps in recognizing the phase of surgery. The proposed approach is tested on Cholec80 dataset. It gives an accuracy of 84.9%.

Keywords: *Workflow Recognition, Tool Presence Detection, Feature Extraction, Robot-Assisted Surgery, Minimally Invasive Surgery*

1. Introduction

Minimally Invasive Surgeries (MIS) and Robot-Assisted Surgeries have gained worldwide attention due to their numerous advantages. Open surgeries have many disadvantages. The cuts made are large. The number of hospital days and the pain is also huge. Minimally Invasive Surgeries are a certain category of surgery where small cuts are made. This automatically reduces the number of hospitalization days. These kinds of surgeries uses an instrument called laparoscope to view the organs inside a patient's body. The resulting videos are showed on a video monitor in real-time. Here comes the digital aspect of surgeries.

There are mainly two concerns in the category of surgeries. One is the tools used. Second is the workflow or sequence of steps involved in the surgery. Since the video monitor provides an accurate vision of surgery, its use can be extended. Due to the complexity of surgeries, the chances of occurrence of faults are really high. Hence automatic analysis and evaluation of surgeries become a necessity. The presence of certain tools and their usage together could accurately determine the phase of the surgery. So, if an automated system is in action, faulty procedures can be easily detected. Apart from this, training clinicians and surgeons is a costly procedure. It's always difficult to assign a surgeon with every trainee. Automated workflow recognition systems could dramatically reduce this cost. This is also advantageous in Robot-assisted surgery systems where surgeries are carried out with the help of a robot.

Most of the feature extraction techniques used for surgical phase recognition is not so effective. Binary tool usage signals, surgical triplets are some common features. But these

features are not learned, they are being found. Because, this features are obtained manually by annotating some dataset. But for large dataset and different problems, this is really hard. If there exists some feature extraction techniques that require less annotations, then it would be very easy. Convolutional Neural Networks are effectively used for feature extraction. We propose a CNN based architecture for feature extraction from surgical videos. This features can help in detecting the presence of tools and the phase of surgery.

Nowadays Convolutional Neural Networks are used in many operational fields. CNNs are so popular due to their generality. Alex Net[22] was one among the first predecessors of CNN history. Several new CNN architectures are evolved followed by AlexNet. EndoNet[11] is one among the latest best CNN architectures for surgical phase recognition. Here, in this paper, a deep learning approach for surgical video feature extraction and tool presence detection is proposed. It uses a popular CNN architecture called VGG16.

The main objective of this paper is to develop an efficient feature extraction approach for workflow recognition in laparoscopic surgeries. The approach incorporates the discriminating feature production capability of deep learning networks.

There is a serious scarcity of data in the medical field. Most of the hospitals does not make their data public due to privacy issues. So, it's better to use transfer learning procedures due to data scarcity.

2. Related Works

2.1. Phase Recognition

M. S. Holden et al. [1] proposed a workflow segmentation algorithm using K-means clustering algorithm and Markov model. To validate the algorithm, Ultrasound-guided epidural procedures and Lumbar Puncture Procedure are used [1]. The approach proposed by H. C. Lin et al.[2] encompasses automatic techniques for surgical gesture detection. This approach uses linear discriminant analysis and Bayes classifier [2]. Real-time workflow identification systems will be an essential element of operating rooms in the future. N. Padoy et al. proposed an approach [3] that makes use of average surgery,a probabilistic surgery model and Hidden Markov model. This approach can be used for phase recognition in any endoscopic surgeries. The workflow detection method [4] proposed by O. Dergachyova et al. Detects surgical workflow with the help of video data and instrument usage signal using Surgical Process Modelling (SPM). Hidden semi-Markov Model (HsMM) is used for modeling the temporal aspect [4]. J. E. Bardram et al. proposed a sensor based approach [5]. All the instruments contain passive RFID tags. Tables contain RFID readers. The third sensor is a palm-based sensor to track the usage of instrument. Here[5] Decision Tree classifier is used. F. Lalys et al. proposed a framework[6] for phase recognition in cataract surgeries. The method [7] proposed by G. Forestier et al. uses decision tree classifier. A model called Workflow-HMM is proposed by N. Padoy et al. in this paper[8]. T. Suzuki et al.[9] proposed an audio recording and analysis system to just detect faults in surgery. Three methods for surgical video classification are proposed [10] by L. Zappella et al. The approach proposed by. P. Twinanda et al.[11] consists of a novel method for phase recognition based on Convolutional Neural Network(CNN). This uses Alexnet as the basis. This enables the automatic learning of features from laparoscopic videos.

2.2. Tool Presence Detection and Tool Detection

Hassan Al Hajj et al. proposed a CNN based approach[12] to detect tools used in cateract surgery by computing the optical flow.[16] is also a CNN based approach with line segment detection and it uses spatio temporal context for tool detection. Amy Jin et al. proposed a region-based CNN method[17] for detecting and localizing tools. Duygu Sarikaya et al. proposed a CNN combined with Region Proposal Network approach[19] for tool localization and detection. [20] uses an ensemble approach of both VGGNet and GoogleNet for multi-label tooll presence detection. Several approaches tries to localize surgical tools. This[13] approach is a Random Forest based one. Combination of CNN and RNN is also used for tool segmentation[14]. [15] uses local appearance of tools and shape of tools to detect surgical tools. Jonas Prellberg and Oliver Kramer proposed a ResNet based approach[18] for multi-label classification of 21 tools used in cataract surgery.

2.3. Convolutional Neural Networks

Nowadays, Convolutional Neural Networks are so trending that almost every computer vision problems are solved with the help of it. From classification to localization and detection, it has wide variety of application areas. The era of CNN began from a basic architecture called LeNet-5[21]. But , its successor AlexNet [22] won the ILSVRC2012 challenge and became number one in solving object detection problems. EndoNet[11] is a modified version of AlexNet which also yielded best results. The training process of CNNs are really complex. It require great computational power like GPU. Also, the training process requires large amount of data and time. The data must be labeled also. There is a difficulty in acquiring the data and annotating the data . In [25], it is clear that we always don't need to train a CNN from scratch. We can use models that are pre-trained for some other tasks via transfer learning [25]. Fine-tuned networks are now used for many applications.

3. Proposed System

3.1. Dataset

There are two main datasets that are commonly used:
1. Cholec80
2. Endo Vis

The Cholec80[23] dataset was created by A. P. Twinanda et al. as a part of their research[11]. This dataset consists of total eighty videos of cholecystectomy surgery . The surgeries were performed by expert surgeons at the University Hospital of Strasbourg,France. We use this dataset for training and validation.

The frame rate of Cholec80 videos is 25 fps. The whole dataset is provided with annotations of phases. It also provides annotations of tools. But the tool presence is binary. That means, if a particular tool is present in a frame then it is marked as zero. Otherwise, marked as one.

The tools used to perform surgeries are shown in Figure 1.

Figure 1. Tools in Cholec80 dataset

3.2. Proposed Methodology

The pipeline of our approach is given in Figure 2. The CNN framework is trained via transfer learning. The output of the network is used for tool presence detection. But for phase recognition, this output is treated as image feature.

The proposed deep learning framework for phase recognition uses a popular Convolutional Neural Network architecture called VGGNet or VGG16[23]. It was the first runner up of ILSVRC challenge in 2014. The main advantage of VGG16 is that it is really deep. VGGNet has five convolutional blocks. Multiple convolutional layers are stacked in each convolutional block. All the kernels used are 3*3 kernels. This makes VGGNet much deeper.

The proposed approach uses fine-tuning approach for phase recognition. Fine-tuning is applicable when the dataset is really small. It is the process of using the same model that is pre-trained on some dataset for another task. The basic idea behind fine-tuning is transfer learning. A model that is trained to identify cars can be used to identify trucks. For accurately estimating surgical phase, features are extracted and inputted to a Support Vector Machine. But we don't dig into it much deeper.

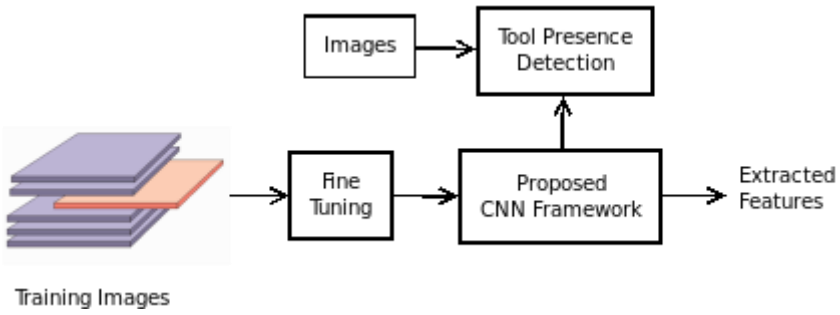

Figure 2: Complete Pipeline of the Proposed Approach

The architecture of proposed CNN framework is shown in Figure 3. We removed the last prediction block of VGG16 and added custom layers that are specific to our task.

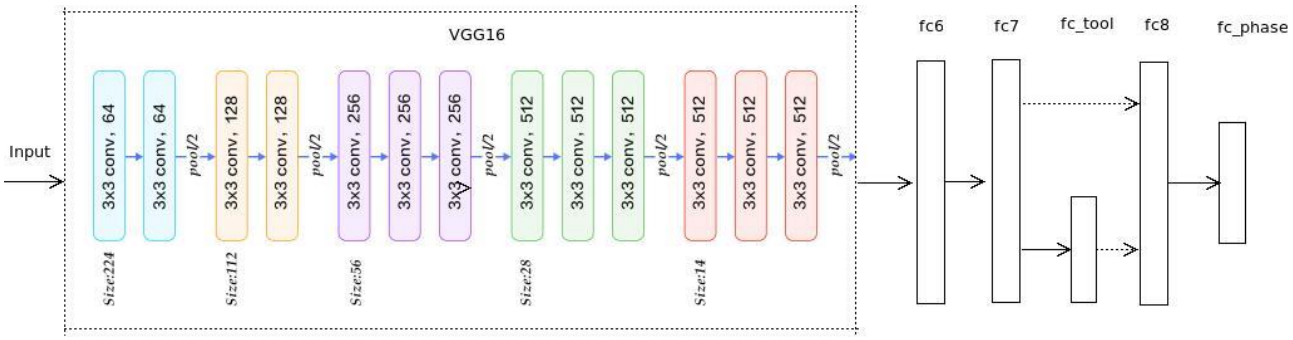

Figure 3: Proposed CNN Framework

The dataset contain videos. We extracted frames from the video and these frames are given as input to the network.

The output of fc_tool gives confidence values that detect the presence of tools. These values are concatenated with output of fc7 at fc8. fc_phase gives the final confidence values for recognizing phase of surgery.

The proposed framework perform two tasks. It detects the presence of tools directly and act as a feature extraction technique for phase recognition.

4. Results and Discussion

The Cholec80 dataset is annotated with tool presence. The proposed architecture is trained with images from Cholec80 dataset. 69044 images are used for training and 17260 images are used for cross validation. The network is trained for 64 epochs. The evolution of loss function is shown in Figure 4.

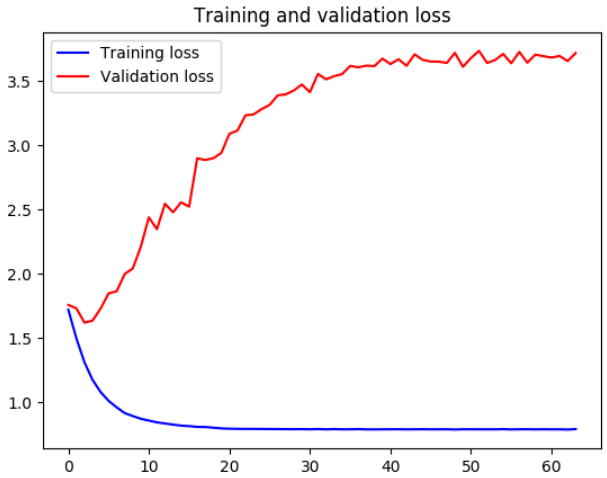

Figure 4: Loss function

X-axis represent epochs and Y-axis represent loss. Convergence can be noticed which indicate that the network is optimized.

The performance of proposed framework in tool presence detection is given in Table 1.

Table 1: Performance of proposed framework in tool presence detection

Sl No.	Tool	Precision	Recall	F1-Score	Support
1	Grasper	0.82384	0.96242	0.88776	60114
2	Bipolar	0.89544	0.54055	0.67414	2614
3	Hook	0.93551	0.71371	0.80970	30305
4	Scissors	0.92593	0.31813	0.47356	943
5	Clipper	0.87156	0.52810	0.65769	1619
6	Irrigator	0.79966	0.53811	0.64332	1758
7	Specimen Bag	0.51151	0.47562	0.49291	841
Accuracy : 0.8493186956433183					

Precision, recall and f1-score is used for performance evaluation. The network accuracy after training is 84.9%.

Since testing the framework with phases doesn't makes much sense, we didn't dig into it much deeper. A single frame is not capable of predicting the phase. This is because, phase itself is a sequence of images. So incorporation of temporal aspect is proposed as future work.

5. Conclusion

In this work, we attempted to find out an efficient feature extraction technique for the phase recognition problem in laparoscopic surgery. The proposed approach also finds the tool presence. The backbone behind the attempt is a Convolutional Neural Network (CNN). The proposed deep learning framework could extract strong features from images. The features are context-dependent. The architecture is fine-tuned with a new dataset. The proposed architecture directly finds the presence of tools and provide features for recognizing phase. The approach yields good accuracy and performance. The obtained features when incorporated with a HMM or its variant could recognize the surgical workflow. Because the introduction of HMM will impose some kind of temporal constraints.

References

[1] *M. S. Holden, T. Ungi, D. Sargent, R. C. McGraw, E. C. Chen, S. Ganapathy, T. M. Peters, and G. Fichtinger, "Feasibility of real-time workflow segmentation for tracked needle interventions," IEEE Transactions on Biomedical Engineering, vol. 61, no. 6,* ***(2014)****,pp. 1720–1728.*

[2] *H. C. Lin, I. Shafran, T. E. Murphy, A. M. Okamura, D. D. Yuh, and G. D.Hager, "Automatic detection and segmentation of robot-assisted surgical motions," in International Conference on Medical Image Computing and Computer-Assisted Intervention,* ***(2005)****,pp. 802–810, Springer.*

[3] *N. Padoy, T. Blum, S.-A. Ahmadi, H. Feussner, M.-O. Berger, and N. Navab, "Statistical modeling and recognition of surgical workflow,"Medical image analysis, vol. 16, no. 3,* ***(2012)****pp. 632–641.*

[4] *O. Dergachyova, D. Bouget, A. Huaulm´e, X. Morandi, and P. Jannin, "Automatic data-driven real-time segmentation and recognition of surgical workflow,"International journal of computer assisted radiology and surgery, vol. 11, no. 6,* ***(2016)****,pp. 1081–1089.*

[5] *J. E. Bardram, A. Doryab, R. M. Jensen, P. M. Lange, K. L. Nielsen, and S. T. Petersen, "Phase recognition during surgical procedures using embedded and body-worn sensors," in Pervasive Computing and Communications (PerCom), 2011 IEEE International Conference on,* ***(2011)****,pp. 45–53,IEEE.*

[6] *F. Lalys, L. Riffaud, D. Bouget, and P. Jannin, "A framework for the recognition of high-level surgical tasks from video images for cataract surgeries," IEEE Transactions on Biomedical Engineering, vol. 59, no. 4,****(2012)****, pp. 966–976.*

[7] *G. Forestier, L. Riffaud, and P. Jannin, "Automatic phase prediction from low-level surgical activities,"International journal of computer assisted radiology and surgery , vol. 10, no. 6,* ***(2015)****,pp. 833 841.*

[8] *N. Padoy, D. Mateus, D. Weinland, M.-O. Berger, and N. Navab, "Workflow monitoring based on 3d motion features," in Computer Vision Workshops (ICCV Workshops), 2009 IEEE 12th International Conference on,* ***(2009)****,pp. 585–592, IEEE.*

[9] T. Suzuki, Y. Sakurai, K. Yoshimitsu, K. Nambu, Y. Muragaki, and H. Iseki, "Intraoperative multichannel audio-visual information recording and automatic surgical phase and incident detection," in Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE, **(2010),**pp. 1190–1193, IEEE.

[10] L. Zappella, B. B´ejar, G. Hager, and R. Vidal, "Surgical gesture classification from video and kinematic data,"Medical image analysis, vol. 17,no. 7, **(2013)**,pp. 732–745.

[11] A. P. Twinanda, S. Shehata, D. Mutter, J. Marescaux, M. De Mathelin, and N. Padoy, "Endonet: A deep architecture for recognition tasks on laparoscopic videos,"IEEE transactions on medical imaging, vol. 36, no. 1,pp. 86–97,**(2017)**.

[12] H. Al Hajj, M. Lamard, K. Charri`ere, B. Cochener, and G. Quellec, "Surgical tool detection in cataract surgery videos through multi-image fusion inside a convolutional neural network," in 2017 39th annual international conference of the IEEE engineering in medicine and biology society (EMBC), pp. 2002–2005,**(2017)**,IEEE.

[13] M. Allan, S. Ourselin, S. Thompson, D. J. Hawkes, J. Kelly, and D. Stoyanov, "Toward detection and localization of instruments in minimally invasive surgery,"IEEE Transactions on Biomedical Engineering, vol. 60, no. 4,pp. 1050–1058, **(2013)**.

[14] M. Attia, M. Hossny, S. Nahavandi, and H. Asadi, "Surgical tool segmentation using a hybrid deep cnn-rnn auto encoder-decoder," in 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC), pp. 3373–3378, **(2017),**IEEE.

[15] D. Bouget, R. Benenson, M. Omran, L. Riffaud, B. Schiele, and P. Jannin,"Detecting surgical tools by modelling local appearance and global shape,"IEEE transactions on medical imaging, vol. 34, no. 12, pp. 2603–2617,**(2015)**.

[16] Z. Chen, Z. Zhao, and X. Cheng, "Surgical instruments tracking based on deep learning with lines detection and spatio-temporal context," in 2017 Chinese Automation Congress (CAC), pp. 2711–2714,**(2017)**,IEEE.

[17] A. Jin, S. Yeung, J. Jopling, J. Krause, D. Azagury, A. Milstein, and L. Fei-Fei, "Tool detection and operative skill assessment in surgical videos using region-based convolutional neural networks," in2018 IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 691–699, **(2018)**,IEEE.

[18] J. Prellberg and O. Kramer, "Multi-label classification of surgical tools with convolutional neural networks," in 2018 International Joint Conference on Neural Networks (IJCNN), pp. 1–8, **(2018)**,IEEE.

[19] D. Sarikaya, J. J. Corso, and K. A. Guru, "Detection and localization of robotic tools in robot-assisted surgery videos using deep neural networks for region proposal and detection,"IEEE transactions on medical imaging, vol. 36, no. 7,pp. 1542–1549, **(2017)**.

[20] S. Wang, A. Raju, and J. Huang, "Deep learning based multi-label classification for surgical tool presence detection in laparoscopic videos," in 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017), pp. 620–623,IEEE, **(2017)**.

[21] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner,et al., "Gradient-based learning applied to document recognition," Proceedings of the IEEE, vol. 86, no. 11,pp. 2278–2324, **(1998)**.

[22] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in Advances in neural information processing systems, pp. 1097–1105, **(2012)**.

[23] Cholec80 Dataset,[Online]. Available:http://camma.u-strasbg.fr/datasets. ,[Accessed: 19-sept-2018].

[24] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition,"arXiv preprint arXiv:1409.1556, **(2014)**.