ISASE 2019

Workflow Recognition from Knee Surgical Videos: Role of Deep Neural Networks

Shoichi NISHIO[1*], Belayat HOSSAIN[1], Manabu NII[1], Takafumi HIRANAKA[2], Syoji KOBASHI[1]

[1] *Graduate School of Engineering, University of Hyogo, 2167 Shosya, Himeji 671-2280, Japan*
[*] *Corresponding author: messi240exe@gmail.com*
[2] *Takatsuki Hospital, 1-3-13 Kosobe, Takatsuki, Osaka 569-1192, Japan*

Abstract: The Orthopedic surgery differs from its counterpart surgery, such as laparoscopic and laparotomy, because of a large variety of surgical techniques. Furthermore, the procedures are complicated, and many types of equipment's have been using in the Orthopedic surgery. So, nurses who deliver surgical instruments to surgeon are supposed to be forced to incur a heavy burden. In our previous work, the navigation system for assisting operating room nurse in Orthopedic surgery – Unicompartmental knee arthroplasty (UKA)- was proposed, and we achieved satisfactory accuracy for operation and out-of-operation phase detection. In this work, we propose method for improvement of recognizing Orthopaedics procedures from video images of UKA and TKA (Total knee arthroplasty) by deep neural network. Firstly, we construct the recognition model and then evaluate the recognition of procedures using convolutional neural network, and finally investigated the role of deep and densely connected neural nets for improving the recognition accuracy. The outcome confirmed the improved recognition rate with deep and dense layers, however, further improvement is yet to discover in future.

Keywords: computer-aided orthopedic surgery, *operating room nurse, surgery video, surgical procedure recognition, deep learning*

1. INTRODUCTION

Knee surgery is popular in clinical to regain the normal functionality of knee. There are some types of knee surgeries, among them knee replacement surgery is an operation to replace damaged knees with artificial knee joints for knee joints damaged by osteoarthritis (OA) or rheumatoid arthritis (RA) [1]. The number of surgeries has increased year by year (the number of domestic cases: more than 80,000) and it is one of major orthopedic surgery (Yano Research Institute Ltd, 2010).

Although total knee replacement arthroplasty (TKA) replaces all knee joint surfaces, however, Unicompartmental knee arthroplasty (UKA) replaces a part of the knee joint surface due to mild damage. The features of these procedures are complicated (roughly 27 procedures) in the procedure, and there are many surgical instruments (about 120 types) to be used, and the operation of assembling the instruments also occurs during the operation. Furthermore, surgical procedures and equipment to be used are different depending on the implant manufacturer and the implant type. Many hospitals use multiple models.

For instrumental nurses who are in charge of multiple kinds of surgical operations, grasping these complicated procedures and equipment to use is a heavy burden [2]. It is concerned that this is a serious incident leading to a surgical mistake, such as surgical time extension, poor prognosis, deterioration in the quality of surgery due to unfamiliar instrumentation, procedures before and after the procedure, instrument errors and so on. Therefore, development of an intraoperative navigation system that supports instruments and nurses in real time is desired.

As a previous research, Twinanda proposed method named EndoNet [3] that using both CNN and Hierarchical HMM is to recognize surgical video images in the field of laparoscopic surgery. However, the Orthopedic surgery differs from its counterpart surgery, such as laparoscopic and laparotomy, because of a large variety of surgical techniques. Furthermore, the procedures are complicated, and many types of equipment's have been using in the Orthopedic surgery. In our previous work [4], we proposed method that how to construct navigation system for operating room nurse in the field of orthopedic surgery and recognition model for recognizing UKA surgical video images using convolutional neural network (CNN) [5].

However, our research has not achieved high accuracy for recognizing surgical image of UKA, and practical method to assist operating room nurse have not established in the field of orthopedic surgery. In this research, we propose a method that more efficient recognition model to recognize UKA and TKA surgical image using deep neural network. and finally investigated the role of deep and densely connected neural nets for improving the recognition accuracy. The outcome confirmed the improved recognition rate with deep and dense layers, however, further improvement is yet to discover in future.

This paper consists of the following chapters. Section 2.1 and 2.2 explains data acquisition and description. Section 2.3 explains the proposed recognition model to identify current surgical procedure. Section 3 explains the experiment results and explanation of the outcome. Finally, the study is concluded with future scope in Section 4.

2. MATERIALS & METHOD

2.1 Data acquisition

With the aim for real time surgical workflows recognition, this study acquired the whole surgery using a video recorder which offers a convenient way to record during Orthopaedics surgery without hampering the surgical task. Thus smart eyeglasses were chosen for recording the surgery as a motion pictures. Basically it's a small wearable electronic device with optical head-mounted display which has many features, including video recording, wireless connectivity, and so on. While it has many industrial application, however, it's also worth for healthcare and clinical application because it lighter than its counterpart, head mounted camera. Furthermore, its capable for bidirectional communication, hands-free operation, and above all augmented reality (AR) in real-time Based on these features, smart glasses are very compatible to record the surgical steps in the surgical navigation systems.

The videos of UKA and TKA surgeries were recorded by smart glasses (InfoLinker, West Unititis Co., Ltd., JAPAN) and archived in the private database of a local hospital. The smart glass was worn by a surgeon while executing the surgery. ZIMMBER BIOMET's implant and surgical instruments were used in the orthopedic surgery (UKA, and TKA). This study was approved by the local Ethics committee of Takatsuki Hospital (Takatsuki, Japan).

2.2 Data annotation

Current study included six UKA and TKA videos. Explanation of the collected raw video images are given in Table 1. The dataset consists of heteronomous videos including different frame rates and sizes. Example of frames of the dominant phases of the UKA surgery are shown in Fig 1. Poor room illumination, tiny and similar surgical instruments and constrained environment make the orthopedic surgical tasks difficult to recognize from the video other than non-orthopedic surgery.

Table 1: Description of Orthopedic surgical video

Video#	Length *[hr:m:s]*	Frame	
		Rate *[f/s]*	**size**
Video 1	0:47:37	24	640×480
Video 2	1:16:21	5	640×480
Video 3	1:27:44	5	320×240
Video 4	1:12:09	5	640×480
Video 5	1:04:55	29	1280×720
Video 6	00:43:35	29	1280×720

The training videos are annotated as follows. Firstly, the surgical phases are manually extracted using a video reader and then annotated accordingly. In this study we considered eleven important surgical phases in UKA and TKA.

2.3 Recognition model

Automatic Orthopaedics surgical procedure recognition model is aimed at recognizing a current procedure from image sequences of the video. We use CNN to extract features from images and classify surgical images. In this work, performance of the model is evaluated for three neural nets- MobileNet [6], VGG16 [7] and Dense Net [8]. For each case, the network is pre-trained in ImageNet images [9]. As a case using VGG16, the one of our CNN model is shown in Fig 2.

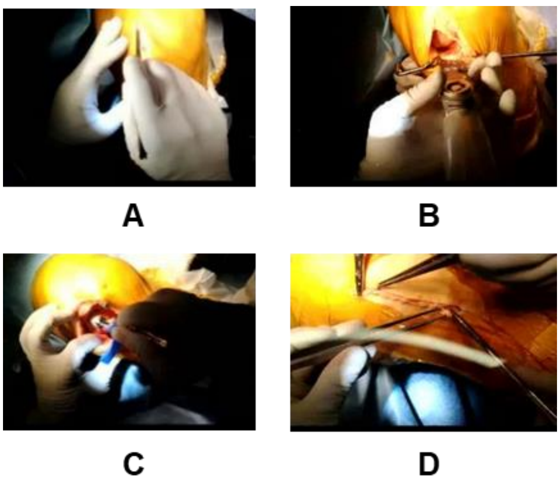

Figure 1: Examples of procedure shown in a frame taken from different procedures in UKA. Procedure- 1 (A), 7 (B), 16 (C), and 27 (D).

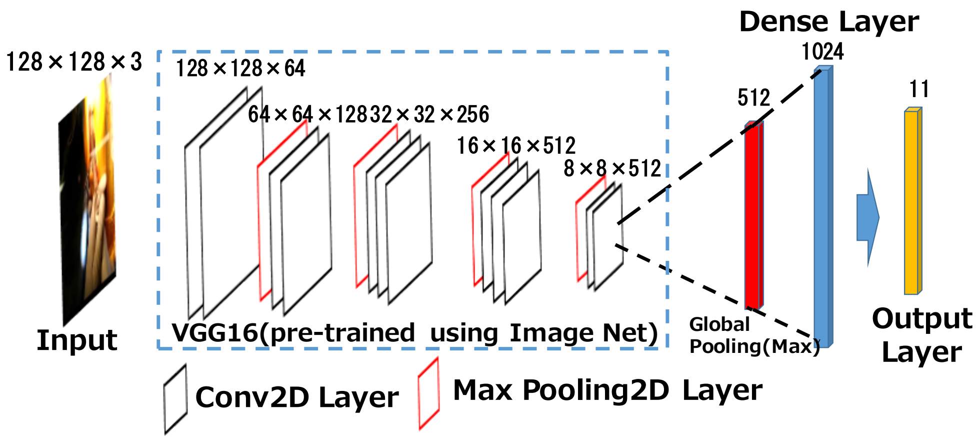

Figure 2: VGG16 based recognition model

Fig 2 model was made by next procedure. First, we cut output layer including fully-connected and Softmax on these model. Second, we added new fully-connected and Softmax layer as output layer on these model. Also, we tried two methods as how to train weights of CNN. As a first method named fine-tune, we train only output layer added by us. As a second method named full-tune, we train weight of the entire model. In addition, we verified the influence on the accuracy of the model by using SVM as a discriminator.

Input single surgical image for our model is 128×128 size and RGB. Image Features has 1024-dimension value was obtained from surgical image by convolution and max pooling. Output vector have the probability of each class was obtained from output layer in Fig.3 using image features. Following this vector, our model determines the most probable class as current procedure.

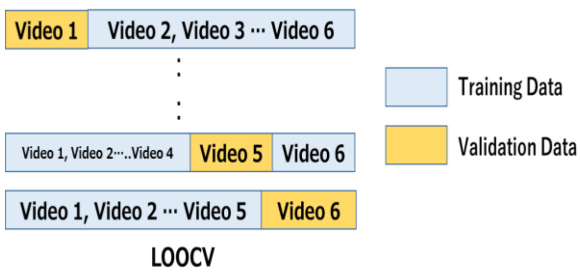

Figure 3: Procedures of cross-validation method

4. EXPERIMEMTAL RESULTS AND DISCUSSION

To evaluate the model, we used the following data shown in Table 2. The verification method was used by the Leave One Out Cross Validation (LOOCV) procedure [10] as shown in the Figure 6. For example, firstly we used dataset of video 2, 3, 4, 5, 6 for training, and video 1 was used for test data. We conducted six tests following the LOOCV procedure until we used entire dataset including all videos for evaluation.

The learning parameters were set using the compile function and fit function of Keras API (Python Library). We used stochastic gradient decent (SGD) as optimizer for training our model. Learning late of SGD was set to 0.0001, and momentum of SGD was 0.90. Loss function is categorical cross entropy in this experiment. Training was conducted with both epochs 10 and 15. We also conducted fine- and full- tuning during model construction for each network. In the output layer, both Softmax and support vector machine (SVM) classifiers were used individually.

The average recognition accuracy result of Mobile Net, VGG16, and DenseNet using LOOCV with Softmax classifier is shown in Table 3. In the following, we considered network trained with epochs 10 because the highest generalization performance was achieved for epochs 10. Furthermore, the full-tuning the network provided better result than fine tuning where all the weights in the network are weighted as ImageNet learned and all of them are re-learned (full-tune) with the data set of Table 2. DenseNet with fine tuning obtained better accuracy, however, VGG16 with full-tuning achieved better accuracy than that of other networks. Full- tuned VGG16 net with Softmax classifier provided 12.06%. 32.11% better average recognition accuracy than that of Mobile-Net and Dense-Net. Therefore, deep neural network could be a good candidate for orthopedic surgical procedure recognition in future. In addition to this, network architecture could be optimized to enhance the recognition accuracy.

Table 2: Dataset to validate the model- total images in each class used during testing.

Video #	Class										
	1	3	6	7	8	10	12	13	16	18	27
1	39	39	39	39	39	39	39	39	26	39	39
2	26	26	13	26	13	26	26	26	26	39	39
3	47	52	0	39	0	0	26	26	52	26	52
4	32	32	0	26	0	0	0	45	0	0	32
5	96	96	78	52	78	39	78	96	78	78	112
6	91	91	91	78	78	78	65	104	91	52	91

Table 3: Average recognition accuracy over all videos during LOOCV validation with Softmax classifier.

CNN model	Epochs -10 [%]		Epochs - 15 [%]	
	Fine-tuning	Full-tuning	Fine-tuning	Full-tuning
MobileNet	22.7	29	22.2	26.5
VGG16	16.7	32.5	17.7	31.3
DenseNet	28.0	24.6	25.6	25.3

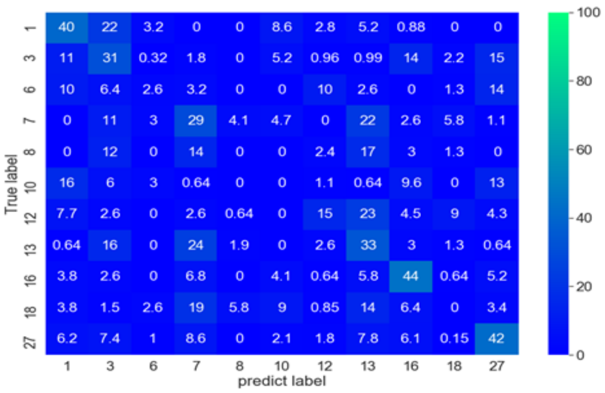

Figure 4: Confusion matrix of the best CNN model (VGG16 + full-tuning).

The confusion matrix related to recognition accuracy of VGG16 (full-tune) is shown in Fig 4. Out of the 11 classes, none of them have good accuracy. Therefore, further work is needed to improve the class accuracy.

Finally, we investigated the performance of the SVM and Softmax classifiers at the output layer of best networks (VGG16). Output layer (Softmax) of the most accurate VGG16 (full-tune) was changed to linear SVM (*l*-SVM), and the same evaluation experiment (including LOOCV procedure) as described above was conducted. Table 4 summarizes the outcome of their comparison. It was found that Softmax classifier outperformed over linear SVM classifier with 7% increases in terms of average recognition accuracy. However, further analysis is need because accuracy deteriorated in some cases for SVM.

Table 4: Average recognition accuracy over all videos during LOOCV validation and epochs-10 with Softmax and SVM classifiers.

Video #	VGG16+ Softmax [%]	VGG16+ *l*-SVM [%]
1	26.5	33.2
2	18.9	23.5
3	45.9	35.3
4	42.4	34.1
5	30.5	21.2
6	30.4	33.3
Average	**32.43%**	**30.10%**

5. CONCLUSION

We proposed automated orthopedic surgical procedure recognition model to recognize current surgical procedure from surgical video acquired from smart glasses. The proposed computer aided navigation system could assist operating room nurses to conceive current surgical procedure, and then to deliver the necessary tools to surgeon during the orthopedic surgery. In this experiment, we investigated that the recognition accuracy for Orthopaedics surgical procedure was improved with full-tuning in deep and densely connected neural network that traditional CNN model.

In future, we need to increase data set for improving our model, to apply data augmentation, and to find optimal parameter in the model.

ACKNOWLEDGMENTS

This work is supported by KAKENHI JSPS Grant-in-Aid grant (18F18377).

REFFERENCES

[1] D. W. Murray *et al.*, "A randomised controlled trial of the clinical effectiveness and cost-effectiveness of different knee prostheses: the Knee Arthroplasty Trial (KAT)," *Health Technology Assessment,* vol. 18, no. 19, 2014.

[2] T. Borus and T. Thornhill, "Unicompartmental Knee Arthroplasty," *JAAOS - Journal of the American Academy of Orthopaedic Surgeons,* vol. 16, no. 1, pp. 9-18, 2008.

[3] 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 Trans Med Imaging,* vol. 36, no. 1, pp. 86-97, Jan 2017.

[4] S. Nishio, M. Hossain, B. Hossain, M. Nii, T. Hiranaka, and S. Kobashi, "Real-Time Orthopedic Surgery Procedure Recognition Method with Video Images from Smart Glasses Using Convolutional Neural Network," presented at the 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC), 2018.

[5] I. G. Y. B. a. A. Courville, *Deep Learning*. MIT Press, 2016.

[6] A. Howard *et al.*, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," *arXiv,* 2017.

[7] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *arXiv e-prints,* Accessed on: September 01, 2014Available: https://ui.adsabs.harvard.edu/¥#abs/2014arXiv1409.1556S

[8] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely Connected Convolutional Networks," *arXiv e-prints,* Accessed on: August 01, 2016Available: https://ui.adsabs.harvard.edu/¥#abs/2016arXiv160806993H

[9] J. Deng, W. Dong, R. Socher, L.-J. Li, L. Kai, and F.-F. Li, "ImageNet: A large-scale hierarchical image database," presented at the 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009.

[10] R. G. Miller, *The jackknife-a review*. Biometrika, 1974.