Bangladesh University of Business and Technology (BUBT)

Course No. : **CSE - 465**

Course Title: **Machine Learning**

Semester: **Summer 2022 [Tri-Semester]**

Survey Paper Name: Hand Frame Extraction in Surgical Video Images Using Convolutional Neural Network

Submitted by –

Name	**ID**	**Task Covered**
Hashibur Rahman Redoy	18192103276	Summary
Hasan Al Mahmud	18192103239	
Tahmid Hossain Rasel	18192103168	Unique Contribution
Masudul Islam Asif	18192103147	Conclusion
Mahbubur Rahman Khan	18192103150	Advantages
Asadul Al Galib	18192103	

Submitted to -

Khan Md. Hasib

Lecturer, Department of CSE, BUBT

Summary of the Paper:

Hand frame extraction is one of the important and crucial tasks to analyze video images of Orthopedics surgery. Although a number of vision-based techniques have been used to interpret hand movements in the field of computer vision, proper hand frame extraction in OS video is still a challenging task in the surgical work atmosphere. Components consisting of metal and plastic are being used to seal the surfaces of the knee assembly bone and form the knee joints. The complicated procedures make difficult to recognize the current progress of the surgery by surgical assistant, nurses, and technicians. Thus, it is desired to develop a computer-aided surgery navigation system, which instructs the progress and the condition, and navigates the operation staffs during the surgical operation. Deep learning in recent years is gaining significant popularity in a wide range of domains such as the recognition of surgical videos which demonstrates the valuable potential for the advancement of contemporary medical care and treatment. More notably, the improvement of the surgical assistant system's real-time environmental detection, the online video recognition technology would in the future be an important aspect of the OR to strengthen the quality of surgery [6]. Hand frame extraction is one of the important procedures to analyze the OS video images. Data augmentation is a method widely used in DL, and it helps to generate the number of samples required. The structure of ResNet-50 used for the hand detection is depicted in Fig.3.

Unique Contribution of the Paper:

In this experiment, they have utilized four UKA videos for training ResNet-50, the remained one UKA video was used for validation. For each class, 1000 images were randomly chosen from the four training videos. And, 333 images were extracted from the validation video for each of hand, no-hand, no-surgical area classes, respectively.

ResNet-50 can perform reasonably well in the case of the UKA surgery hand detection and classification along with the other surgical environments. The ResNet-50 could classify almost all the classes available in UKA dataset. However, the model got confused whilst classifying the no-hand and non-surgical area classes, because of some of the identical regions between both the two classes.

They presented a method to detect different surgical environments from UKA surgery video images for the purpose of hand detection. The images from the videos were extracted in JPEG format and divided into three classes (hand, no-hand, and non-surgical area). They applied ResNet with 50 layers on our dataset consisting of 1000 images in class. Results demonstrate that the models' performance with an overall accuracy of 96%. Moreover, the other performance measures for each class were calculated. It was found that the hand class has the highest F1-score of 0.97, no-hand class holds the highest recall of 0.99, and the nonsurgical area class with the highest precision of 1.00. Our future work includes extending this approach to detect hand based on the surgical phase using deep learning. The region of hand will be segmented for a few specific surgery phase hence by preparing the dataset the model will be evaluated based on the desired algorithm.

How the proposed model works in the paper:

Advantages of the paper:

Surgical skill has a significant impact on the long-term survival of patients and a higher skill level is associated with fewer postoperative complications and better patient outcome The usage of the JIGSAWS data set in studies on the automatic assessment of technical and surgical skills emphasizes the importance of annotated data sets for the improvement of surgical training through personalized feedback However, it does not contain real patient data It enables the documentation of training success as well as the analysis of personal and structural deficits with the intention of specifically addressing them The collected data were integrated into the automated, objective evaluation of surgical skills

Disadvantages of the paper:

Conclusion:

In this work, presented a method to detect different surgical environments from UKA surgery video images for the purpose of hand detection. The images from the videos were extracted in JPEG format and divided into three classes (hand, no-hand, and non-surgical area). We applied ResNet with 50 layers on dataset consisting of 1000 images in class. Results demonstrate that the models' performance with an overall accuracy of 96%. Moreover, the other performance measures for each class were calculated. It was found that the hand class has the highest F1-score of 0.97,no-hand class holds the highest recall of 0.99, and the non-surgical area class with the highest precision of 1.00. Future work includes extending this approach todetect hand based on the surgical phase using deep learning. The region of hand will be segmented for a few specific surgery phase hence by preparing the dataset the mode l will be evaluated based on the desired algorithm.