N. Ding *, N. A. Jalal, T. A. Alshirbaji and K. Möller

# The evaluation of synthetic datasets on training AlexNet for surgical tool detection

**Abstract:** Surgical tool recognition is a key task to analyze surgical workflow, in order to improve the efficiency and safety of laparoscopic surgeries. The laparoscopic videos are important sources to conduct this task, However, there are some challenges to analyze these videos. Focus on the imbalanced dataset problem, data augmentation method based on generate different synthetic datasets and evaluate their performance training on a convolutional neural network model are investigated in this research. The results show the effect on the model with different background patterns. A better performance was achieved when the model was trained by a structure background dataset. Further research will be needed to understand why the original background patterns support the correct classification. It is assumed that this is an overlearning effect, that will not hold if other procedures were included into the test set.

**Keywords:** surgical tool recognition, convolutional neural network.

https://doi.org/10.1515/cdbme-2020-3082

## 1 Introduction

Laparoscopic videos contain valuable information of minimally invasive surgeries, such as surgical tools, surgical actions and tissues [1]. Surgical tool recognition based on the analysis of these laparoscopic videos have gained increasing attention by researchers due to its importance to better control surgical workflow. Recognizing surgical phases is an essential component to develop context-aware system (CAS), in order to optimize operating room procedures, support surgical teams, increase the efficiency of surgeries. It would also be possible to automate the indexing of surgical video databases. Such context-aware systems could also be used to alert the clinicians to probable upcoming complications [2].

However, there are several difficulties for detecting surgical tool presence. For instance, it is a multi-label classification task, several tools can be used simultaneously. In cholecystectomy surgery videos, four tools can appear at the same time, which makes the task challenging. Also, the camera is not static during surgery, resulting in motion blur and high variability of the observed scene. The complexity of endoscopic images, specular reflection, blurs due to smoke, or partial occlusion of the tool by an anatomical structure or blood, all these situations render tool recognition a difficult task. In addition, some tools appear more frequently than others, causing the imbalanced dataset problem [1,2]. Nevertheless, it greatly affects the training process of CNN model. Related work, that focus on this imbalanced dataset problem, used different approaches to decrease the influence of tool usage frequency, such as resampling to get a balanced dataset and loss-sensitive learning [1].

In this work, we take a first step toward generating synthetic data that can be used to augment available datasets in order to improve tool presence detection using CNNs. The influence of the background to the training efficiency is explored. Therefore, three artificial datasets were generated by substituting the image background by three different patterns, one structured and two unstructured backgrounds, original-backgrounds, uniform-backgrounds and random-backgrounds. An evaluation of these datasets in terms of their effectivity to train a CNN model for tool detection was made.

## 2 Method

Three balanced datasets were generated by applying image transformations and substituting image backgrounds of real images from the Cholec80 dataset [2]. It is a large dataset of cholecystectomy videos containing videos of 80 surgical procedures recorded at the University Hospital of Strasbourg. The videos are captured at 25 fps and down sampled to 1 fps

***Corresponding author: N.Ding:** Institute of Technical Medicine, Jakob-Kienzle-Strasse 17, VS-Schwenningen, Germany, e-mail: ning.ding@hs-furtwangen.de
**N. A. Jalal, T. A. Alshirbaji, K. Möller:** Institute of Technical Medicine, VS-Schwenningen, Germany

for processing. The whole dataset is manually labeled with the phase and tool presence annotations. At first, for every surgical tool, images were cropped from a real video frame to remove the background. Fifty images of each tool were generated from randomly chosen real laparoscopic images. Then, tool images were augmented using 2 types of image-based augmentation methods including image rotation in five angels 0°, 45°, 90°, 135°, 180° and image translation by five vectors in both x and y axes. Finally, three artificial background patterns, monochrome, random and original backgrounds, were employed to acquire three different datasets that are uniform-, random-, and original-background datasets respectively. First we substitute the cropped tool image background with 5 different backgrounds for each dataset, then increasing the backgrounds number to 8 and 15 to investigate if the increase of the size influences the learning process. [4]

In the uniform-background dataset, all background pixels have the same value, while backgrounds in the random-background dataset were generated randomly based on the histogram distribution of a real laparoscopic image. Besides synthetic backgrounds, original images not containing any surgical tool were paired with tool images to produce original-background dataset. The validation dataset used to evaluate the performance of these training sets are 350 real images from which the tools were cropped.

**Figure 1:** Results of training the model with different training sets.

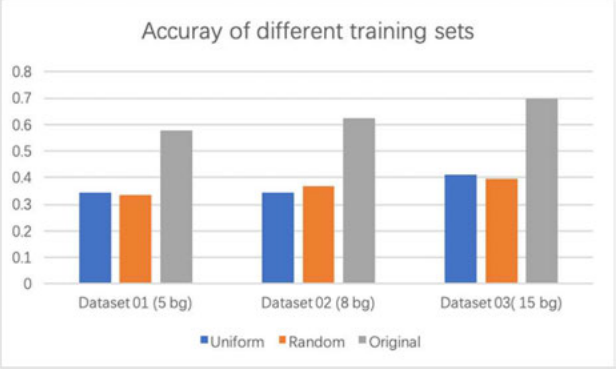



**Table 1:** Another evaluation metric on dataset 01(5 backgrounds).

| Dataset | Avg. Precision | Avg. Recall | Accuracy |
|---|---|---|---|
| uniform | 45.4% | 34.6% | 34.6% |
| random | 28.3% | 33.7% | 33.7% |
| original | 67.8% | 57.7% | 57.7% |

# 3 Results and Discussion

These synthetic datasets are used to train the same model called AlexNet [3]. It is a pre-trained model which contains five convolutional layers and three fully-connected layers. The final fully-connected layer is replaced to adjust to the cholecystectomy data, which shows seven different classes i.e. seven surgical tools are to be recognized. In all trials, the initial weights are controlled i.e. the runs can be repeated exactly.

The evaluation of synthetic datasets started with 5 backgrounds. Each CNN was trained for 5 epochs, which was sufficient for convergence. After training with each dataset and evaluation on the same validation set, the different classification accuracy in the results indicate that the model learnt to recognize (some of) the tool objects, but the learning is also strongly affected by the background patterns. By comparing the results between different background patterns, the original (structured) background datasets have shown better performance than the other two background datasets. (see Fig.1).

Further insight could be gained by evaluation of the confusion matrices for each trainings result. In Fig. 2 a confusion matrix is showed for the uniform background with 5 different background patterns. It exemplarily shows how the models trained with the uniform or random background datasets, misclassify some tools, which seem to be more difficult to distinguish. These misclassifications are responsible for the lower precision and recall as shown in Table 1. Some of those misclassifications could be avoided when training with the structured background (Table 1 & Fig.3). These performances indicate that when training the CNN model, the structured background influences the classification result, which means that the classification of the tools is not solely based on properties of the tool occurrence in the images, but in addition on some background features. This is clearly unwanted as it will limit generalization to other operations with a different background. In the worst case the background could dominate the classification and may lead to critical misclassifications.

Further tests were conducted by increasing the size of each dataset for training. We increase the background number to 8 and 15 different patterns and repeated the training process. Figure 1 depicts the results of training the model according to different sizes, for random and uniform training dataset, the increment has not much effect on the performance, however, the training with the original background datasets show slight improvement after size augmentation.

**Figure 2:** Confusion matrix of training the model with uniform background dataset with 5 backgrounds.

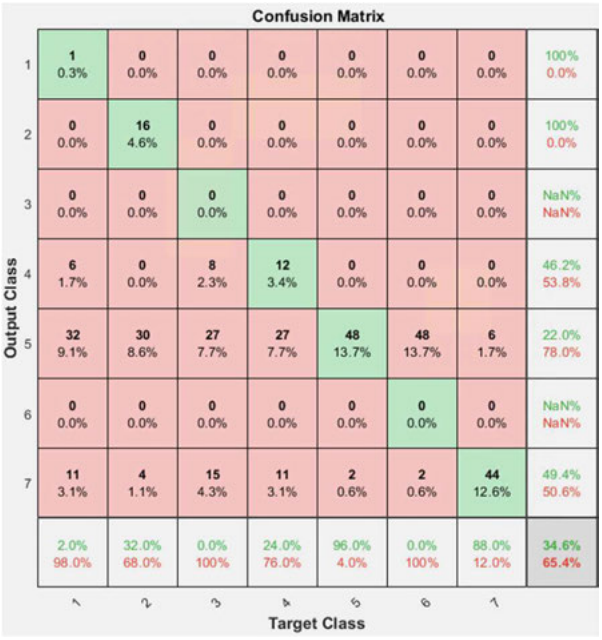
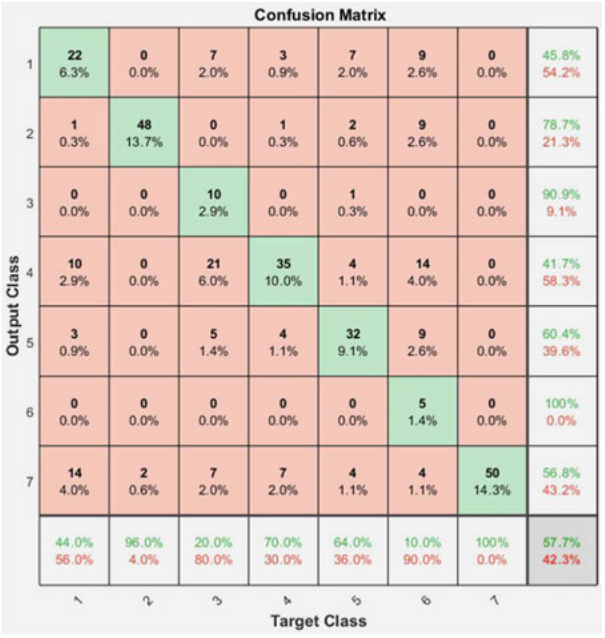



**Figure 3:** Confusion matrix of training the model with original background dataset with 5 backgrounds.




# 4 Conclusion

This research shows that different background patterns of synthetic datasets affect training of a CNN model, and especially the structured background has influence on the classification. This demonstrates a typical problem for machine learning applications: a better performance index is not identical to a better classification performance based on the objects properties. The large number of parameters (access degree of free parameters) in the network architecture carries the risk of unwanted performance.

The limitation of this experiment is that the test set is manually balanced, which caused the evaluation metric recall and accuracy are at same value.

Further tests will include a further class, which has no tool appearance, a situation that occurs in real cholecystectomy surgery video frames. In addition, it will be checked why some tools are so difficult to classify from the image alone. The planned next steps will focus on the original background patterns, generate more data for training in order to get better performance, then incorporate with the real images to ameliorate the imbalance dataset problem.

### Author Statement

Research funding: This work was supported by the German Federal Ministry of Research and Education (BMBF under grant CoHMed/IntelliMed grant no. 13FH5I01IA and 13FH5I05IA). Authors state no conflict of interest. Informed consent: Informed consent has been obtained from all individuals included in this study. Ethical approval: The research related to human use complies with all the relevant national regulations, institutional policies and was performed in accordance with the tenets of the Helsinki Declaration, and has been approved by the authors' institutional review board or equivalent committee.

## References

[1] Alshirbaji, T.A., Jalal, N.A. and Möller, K., 2018. *Surgical Tool Classification in Laparoscopic Videos Using Convolutional Neural Network*. Current Directions in Biomedical Engineering, 4(1), pp.407-410.

[2] Twinanda, A.P., Shehata, S., Mutter, D., Marescaux, J., De Mathelin, M. and Padoy, N., 2016. *Endonet: a deep architecture for recognition tasks on laparoscopic videos*. IEEE transactions on medical imaging, 36(1), pp.86-97.

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

[4] T. Abdulbaki Alshirbaji*, N. Ding, N. A. Jalal, and K. Möller ,2020.*The Effect of Background Pattern on Training a Deep Convolutional Neural Network for Surgical Tool Detection.* Paper ID: 024, DOI: 10.18416/AUTOMED.2020.