The effect of background pattern on training a deep convolutional neural network for surgical tool detection

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

Institute of Technical Medicine, Furtwangen University, Villingen-Schwenningen, Germany
* *Corresponding author, email: abd@hs-furtwangen.de*

Abstract: Surgical tool presence detection in laparoscopic videos plays a key role in developing context-aware systems (CASs). These systems are designed to support surgical team inside operating rooms and increase the efficiency of surgical workflow. Convolutional neural networks (CNNs) have shown robust performance in detecting surgical tools in laparoscopic images. However, imbalanced data sets are still influencing the training process of the CNN models. In this work, data augmentation methods based on generated artificial images as training patterns by substituting the image background by uniform, random or original background patterns are investigated. First experimental results show different effects on the training process. Easily, an improvement of 10% in classification accuracy could be achieved when the network was trained on augmented data.

I. Introduction

Surgical tool recognition in laparoscopic videos is one of the key tasks that has attracted extra attention by researchers when starting to analyze surgical workflow. It has potential applications such as recognizing surgical phases which is a core problem for developing a context-aware system (CAS). However, surgical tool presence detection has several characteristic difficulties that make it a challenging task. For instance, it is a multi-label classification task where several tools can be used simultaneously. In addition, some tools appear more frequently than others, so that the imbalanced dataset problem arises.

Early work on detecting surgical tools employed radio frequency identification (RFID) markers that were directly attached to tools [1]. Other studies, such as [2,3], suggested methods based on handcrafted visual features and classifiers to detect tools in images. In recent years, the large majority of research adopts different classification approaches based on deep learning, namely convolutional neural networks (CNNs), to automatically learn visual features from laparoscopic images and perform the tool classification task [4-6]. For example, in [4] Twinanda et al. proposed a CNN structure called 'EndoNet' that performs tool presence detection and phase recognition.

Existing works explored the imbalanced data problem [6]. Nevertheless, it is still unsolved and greatly affects the training process of CNN models. In this work, by focusing on this problem, 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. Therefore, three artificial datasets were generated by substituting the image background by three different patterns. These background patterns are uniform-backgrounds, original-backgrounds and random-backgrounds. An evaluation of these datasets in terms of their effectivity to train a CNN model for tool detection was made.

II. Material and methods

Three balanced datasets were generated by applying image transformations and substituting image background of real images from the Cholec80 dataset [4]. At first, surgical tool image was cropped from a real image 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^o, 45^o, 90^o, 135^o, 180^o 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.

In the uniform-background dataset, all background pixels have 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. A cross-dataset validation was performed by training AlexNet model [7] on different input combinations.

III. Results and discussion

AlexNet model [7] was fine-tuned using different input combinations (i.e. cross-dataset validation) to investigate improving model generalizability using different artificial background patterns. Fig. 1 shows a summary of results.

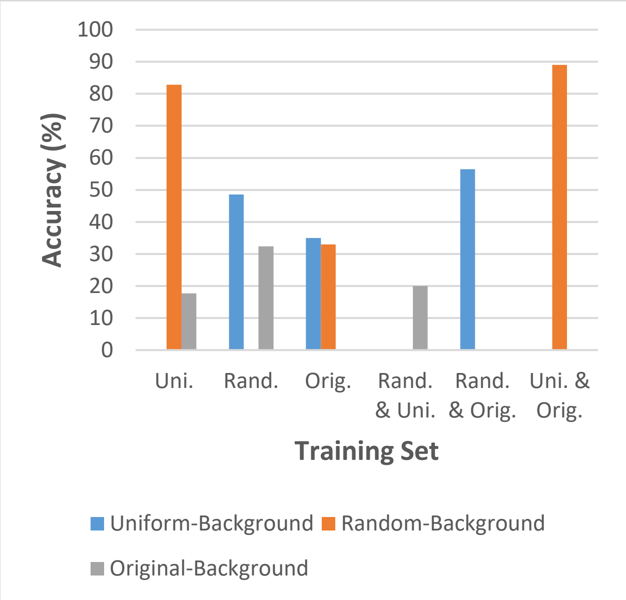

Figure 1: Classification accuracy of cross-dataset evaluation

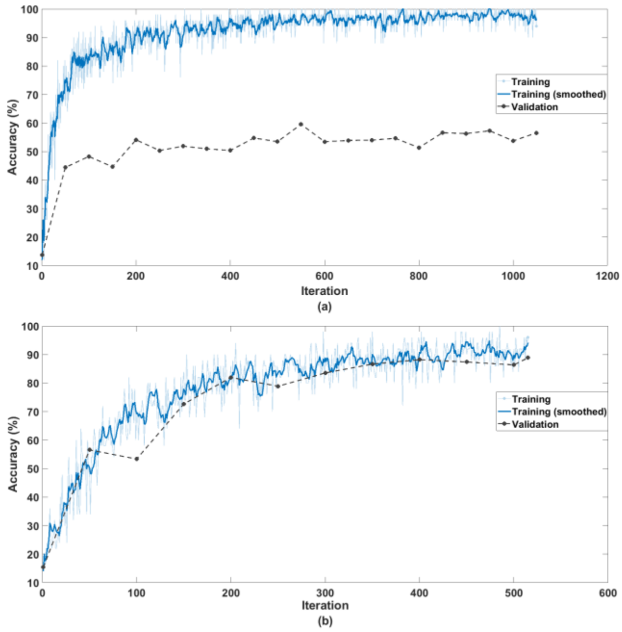

Figure 2: training accuracy of the network. (a) the training set is a combination of rand. and orig. datasets while the validation set is uni. dataset. (b) the training set is a combination of uni. and orig. datasets while the validation set is rand. dataset

It appears that despite having same tool objects in the three datasets, cross-dataset validation, training on one or two datasets while testing on others, gave different classification accuracies (see Fig. 2). These results indicate that visual features learnt by the CNN model are still related to the image background. Interestingly, the CNN model was able to focus on parts of images containing target objects, i.e. surgical tools, when it was trained on uniform-background dataset rather than random-background dataset. Nevertheless, the evaluation against original-background dataset was too low in both cases, that indicates just little generalization capability of the model was obtained, and a strong dependence on background patterns appears to be occurring.

Further testing and analysis were conducted by incorporating real images from the Cholec80 dataset [4]. Here the CNN model was trained using different combinations of real and artificial images, namely: group 1: real images dataset; group 2: random-background, original-background and real images training set; and group 3: uniform-background, original-background and real images training set, and tested on real images taken from the Cholec80 dataset. Table 1 shows results of testing the network on real images according to different training sets. Experimental results emphasize that an improvement of 10% was obtained when group 1 was augmented with artificial data.

Table 1: Classification accuracy with different combinations of real and artificial data.

Training set	**Classification accuracy (%)**
Group 1	61%
Group 2	71%
Group 3	69%

IV. Conclusions

This work shows that different background patterns could affect training CNNs. Further work is needed to incorporate more random background patterns and more tool images.

AUTHOR'S STATEMENT

Research funding: This work was supported by the German Federal Ministry of Research and Education (BMBF under grant CoHMed/IntelliMed grant no. 03FH5I01IA). Conflict of interest: Authors state no conflict of interest. Informed consent: Informed consent is not applicable. Ethical approval: The research is not related to either human or animals use.

REFERENCES

[1] Kranzfelder, M., Schneider, A., Fiolka, A., Schwan, E., Gillen, S., Wilhelm, D., Schirren, R., Reiser, S., Jensen, B. and Feussner, H., 2013. *Real-time instrument detection in minimally invasive surgery using radiofrequency identification technology*. journal of surgical research, 185(2), pp.704-710.

[2] Doignon, C., Nageotte, F. and De Mathelin, M., 2006. *Segmentation and guidance of multiple rigid objects for intra-operative endoscopic vision*. In Dynamical Vision (pp. 314-327). Springer, Berlin, Heidelberg.

[3] Bodenstedt, S., Ohnemus, A., Katic, D., Wekerle, A.L., Wagner, M., Kenngott, H., Müller-Stich, B., Dillmann, R. and Speidel, S., 2018. *Real-time image-based instrument classification for laparoscopic surgery*. arXiv preprint arXiv:1808.00178.

[4] 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.

[5] Wang, S., Xu, Z., Yan, C. and Huang, J., 2019, June. *Graph Convolutional Nets for Tool Presence Detection in Surgical Videos*. In International Conference on Information Processing in Medical Imaging (pp. 467-478). Springer, Cham.

[6] 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.

[7] 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).