Proceedings on Automation in Medical Engineering
Proc AUTOMED, Vol 2, No 1 (2023): Article Ref 755
www.journals.infinite-science.de/automed/article/view/755

Infinite Science Publishing

The Image flip effect on a CNN model classification

N. Ding[1*], K. Möller [1]

[1] *Institute of Technical Medicine, Furtwangen University, Villingen-schwenningen, Germany*
[*] *Corresponding author, email: din@hs-furtwangen.de*

Abstract: Convolutional neural networks (CNNs) have been successful applied to many fields such as image processing. The CNNs can automatically extract image features and make corresponding predictions. However, some minor changes on the images can lead the neural network to wrong predictions. In this paper, we will use image flip technique to evaluate a CNN performance. The experiments show the model are sensitive to a particular flip direction. Additional training with the preprocessing technique have been proved helpful in improving the robustness of the model.

I. Introduction

Surgical tool recognition is one of the medical application of convolutional neural networks (CNNs), intending to build a context aware system to automatically detect the surgery process [1]. The robustness of the model is one of the most important considerations before it launched in the medical system. Hence, it is necessary to evaluate the model robustness thoroughly. Image flip is a simple image pre-processing technique, to generate images with no different to a legitimate image, while these invisible modifications can also influence the CNNs classification performance. In this paper, we will use image flipping to evaluate a trained convolutional neural network.

II. Material and methods

In our experiments we use transfer learning to fine-tune the convolutional neural network model AlexNet [2] for surgical tool classification. The image dataset is Cholec80 containing 80 cholecystectomy videos, including 7 different tools [3]. We extracted 80,190 images which contain only one kind of tool, from these images 25,000 were used to train the model [4], the rest 55,190 images were used as the test set to evaluate the classification performance. The training state of the CNN model was recorded when the training accuracy reached to: 75%, 85%, 95%, and 99% and maintain 10 iterations, these states were named as model 75, model 85, model 95, model 99.

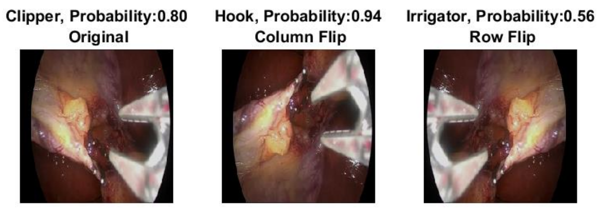

Figure 1: The original image x correctly classified as the surgical tool Clipper. With column-wise flip and row-wise flip, the image was recognized as other tools.

The test set was used to create other two test sets by simply flipping the images column-wise and row-wise. On the second part of the experiment, we use the image-flip technique to augment the training set with column-wise flipping and row-wise flipping images, each flip technique has 50% probability, the training states were recorded at the same training states: 75%, 85%, 95%, and 99%, to compare with the models which trained only with the original images.

III. Results and discussion

In this experiments we evaluate the model performance related to the image flip effect at different training states. The classification performance was evaluated on the 3 test sets (original, column-wise flipped, row-wise flipped). Table 1 shows the classification accuracy were decreased on the flipped images, especially on the row-flipped images. The reason could be the property of Cholec80 dataset. The medical tools used in cholecystectomy were frequently coming through the right side, after the image flipping row-wise the tool appeared on the left side of the image, and the model could not be able to recognize it. It proves that AlexNet model are sensitive to the object positions.

Table 1: The classification accuracy on the original image and column-wise flip and row-wise flip of the test set.

Model	Original set	Column-wise flip	Row-wise flip
75%	81.1%	78.6%	58.3%
85%	90.2%	84%	52.1%
95%	94.0%	86.9%	58.1%
99%	95.2%	86.4%	62.6%

Further evaluation can be measured by F1-score, especially for different classes. Figure 2 shows the flip effect on the particular medical tools at different states, the classification performances were decreased especially for surgery tool scissors, clipper, irrigator and specimen bag, the f1-score values are below 50% after the image-flipping.

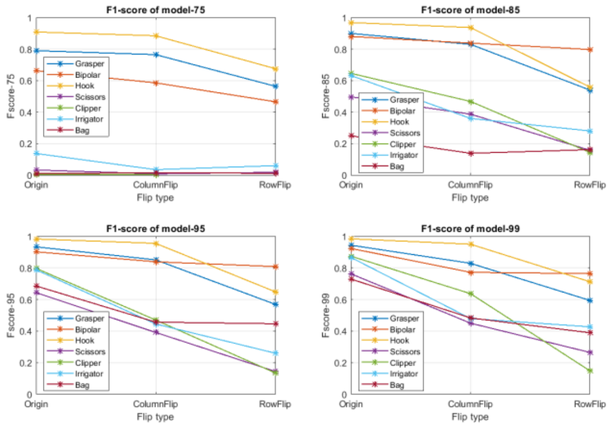

Figure 2: The classification performance of four models on the original test set, the column-wise flipped test set, and the row-wise flipped test set. The classification performance of 7 different surgical tools represents by F1-score, which is the harmonic mean of precision and recall.

In figure 3, the 1st row shows the four models classification performance on the test set, clearly the performance are worse on the test set with images row-wise flipped. Meanwhile, the performances are better on the original images and column-wise flipped images, and the accuracy were improved constantly with better training states.

In the 2nd row, the four models which trained with additional pre-processing, the classification on the same 3 test sets have gained the resistance ability to the image flip effect, now the performances were more stable and improved gradually with the training states.

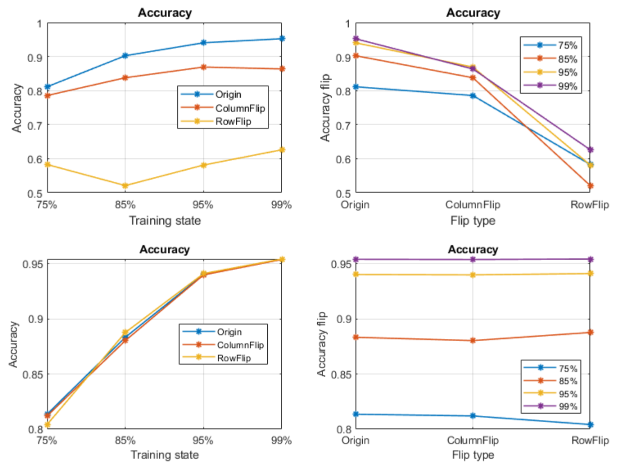

Figure 3: The 1st row showing the classification performance of four models on the original test set, the column-wise flipped test set, and the row-wise flipped test set. The 2nd row showing the classification performance of four models trained with additional pre-processing images on original test set, column-wise flipped test set, and row-wise flipped test set.

Figure 4 also indicates that when the model trained with additional image-flip pre-processing, the classification performances on specific surgical tools were no longer influenced by the image-flipping.

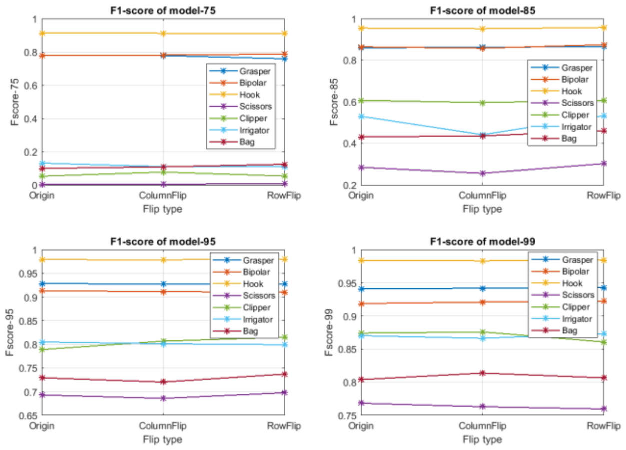

Figure 4: After additional training with image-flip, the F1-score of 7 medical tools on the original test set, the column-wise flipped test set, and the row-wise flipped test set.

IV. Conclusions

In this research, we evaluate a CNN model performance on the image flipping. Additional training with pre-processing the images with the flipping technique can help the model better performed on classifying flipped images.

In future work, we will use the same approach to evaluate other CNN models to compare its robustness or its resistance to image flip.

Author's statement

Research funding: This work was supported by the German Federal Ministry of Research and Education (BMBF under grant CoHMed/PersonaMed KFZ 13FH5I06IA). Conflict of interest: 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., Ding, N., Jalal, N. A., & Möller, K. (2020). The effect of background pattern on training a deep convolutional neural network for surgical tool detection. Proceedings on Automation in Medical Engineering, 1(1), 024-024.

[2] 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).

[3] 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.

[4] Ding, Ning, and Knut Möller. "Generating adversarial images to monitor the training state of a CNN model." Current Directions in Biomedical Engineering 7.2 (2021): 303-306.