Surgical Tool Annotation in Cataract Surgery Videos

Xiaowei Hu and Pheng-Ann Heng

Dept. of Computer Science and Engineering, The Chinese University of Hong Kong, Hong Kong, China

1 Introduction

Automatically annotating surgical tools in surgery videos is critical to analyze the surgical workflows. The surgical workflow analysis has lots of applications including surgical report generation, surgical training and real-time decision support. The goal of surgical tool annotation is to recognize which kind(s) of tool is/are used during the surgery rather than to precisely locate tools in images. This task is challenging for the following reasons. First, two or more kinds of surgical tools may appear in the same images and sometimes are overlapped with each other. Second, usually there is only an image-level label provided for each image and a lack of bounding boxes for individual surgical tools. In principle, this is a multi-label classification task as we need to describe each image using multiple labels and these labels are independent. In cataract surgical videos, the aim is to remove the lens and put artificial lens. During the surgery, the eyes of patients are under the microscope that is video-record.

In this report, we develop a deep learning based framework to annotate surgical tools. This framework takes the solely frame of the surgical video as the input and outputs the label for the current frame. We take two widely used networks as the basic networks [1, 3] to extract the features of the input images and use the cross-entropy loss to train these two networks. Finally, a gate function [2] is used to combine the results of these two networks. Experiments on validation set of the cataract surgery dataset show that our method is able to achieve over 98% annotation performance.

2 Experiments

2.1 Dataset and Evaluation Metric

The cataract surgical dataset consists of 50 videos on cataract surgeries[1]. The frames of these videos are 1920×1080 and the frame rate is around 30 frames per second. The tool is considered to be in use in cataract surgery when it is contacted with the eyeball, which is labeled by two non-M.D. experts independently. When two experts disagree with each other, the label is 0.5, otherwise the label is 0

[1] https://cataracts.grand-challenge.org/

(not used) or 1 (used). There are twenty-one kinds of surgical tools used in these 50 videos in total. The dataset is divided into a training set (25 videos) and a testing set (25 videos). Moreover, the 21 kinds of tools are all appeared in the training set and the testing set. Since we don't have the labels of the testing set, we further divide the training set into sub-training set (the first 20 videos) and validation set (the left 5 videos).

The performance of surgical tool annotation method is evaluated by the area under the ROC curve.

2.2 Implementation Details

Our framework is built on the ResNet101 and DenseNet169 with 101 layers and 169 layers respectively. We change the channel numbers of the final layers on these two networks to 21 that is the same as the number of categories of surgical tools. The input images are resized into 224×224. These two networks are trained independently and the results of these two network are merged by the gate function [2].

During the training process, we ignore the frame labeled as 0.5 and use cross-entropy loss to train our deep networks. To accelerate the training process and reduce over-fitting risk, we initialize the parameters using the well-trained weights on ImageNet classification task. And stochastic gradient descent is applied to optimize the networks with the momentum of 0.9 and the weight decay of 0.0005. For training the sub-training set, we set the learning rate as 0.001, reduce it by 0.1 after 6k iterations, and stop learning after 11k iterations. For training the whole training set, we set the learning rate as 0.001, reduce it by 0.1 after 7.5k iterations, and stop learning after 14k iterations. We use a batch size of 16 for ResNet101 and a batch size of 14 for DenseNet169. In addition, training images are randomly flipped for data argumentation.

2.3 Evaluation on Validation Set

Table 1 shows the results on validation set. Only 16 kinds of tools are evaluated, since the left 5 kinds of tools are not simultaneously appeared in the sub-training set and the validation set. From the Table 1, we can find that almost all the surgical tools can be annotated accurately by these two networks. When we fuse the scores of these two networks, we can further improve the results and the score for each kind of tool is over 0.9.

3 Conclusion

This report presents a framework for surgical tool annotation by harvesting gate function to fuse the results of two basic deep neural networks. we test our framework on validation set, and show our framework achieves over 98% performance on average.

Table 1. Evaluation on validation set.

Method	Final score	ResNet101	DenseNet169
Average score	**0.9812**	0.9744	0.9703
Charleux cannula	0.9622	0.9589	0.9483
Hydrodissection cannula	0.9958	0.9900	0.9961
Rycroft cannula	0.9814	0.9802	0.9702
Viscoelastic cannula	0.9387	0.8917	0.9577
Cotton	0.9596	0.9716	0.8773
Capsulorhexis cystotome	0.9993	0.9982	0.9997
Bonn forceps	0.9897	0.9880	0.9865
Capsulorhexis forceps	0.9721	0.9492	0.9732
Needle holder	0.9933	0.9951	0.9739
Irrigation/aspiration handpiece	0.9972	0.9959	0.9956
Phacoemulsifier handpiece	0.9996	0.9993	0.9993
Implant injector	0.9961	0.9916	0.9967
Primary incision knife	0.9987	0.9976	0.9983
Secondary incision knife	0.9988	0.9979	0.9987
Micromanipulator	0.9855	0.9735	0.9886
Suture needle	0.9310	0.9120	0.8647

References

1. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: CVPR. pp. 770–778 (2016)
2. Hu, X., Yu, L., Chen, H., Qin, J., Heng, P.A.: Agnet: Attention-guided network for surgical tool presence detection. In: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, pp. 186–194 (2017)
3. Huang, G., Liu, Z., Weinberger, K.Q., van der Maaten, L.: Densely connected convolutional networks. In: CVPR. pp. 2261–2269 (2017)