Public Health and Informatics J. Mantas et al. (Eds.) © 2021 European Federation for Medical Informatics (EFMI) and IOS Press. This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0). doi:10.3233/SHTI210206

Semi-Automated Machine Learning Video Annotation for Gastroenterologists

Adrian KRENZER^{a,1}, Kevin MAKOWSKI^a, Amar HEKALO^a and Frank PUPPE^a ^a Julius-Maximilian University of Würzburg, Germany

Abstract. A semi-automatic tool for fast and accurate annotation of endoscopic videos utilizing trained object detection models is presented. A novel workflow is implemented and the preliminary results suggest that the annotation process is nearly twice as fast with our novel tool compared to the current state of the art.

Keywords. machine learning, deep learning, video annotation tool, endoscopy

1. Introduction

Recently machine learning started to play an important role in the domain of medical analysis, classification and disease prevention [1]. Most supervised machine learning algorithms need lots of high-quality data. The annotation process to acquire this data is very costly and labor-intensive, especially if domain experts are involved. Therefore, a tool for an efficient annotation process is presented, which reuses routinely made pathologic snapshots from an endoscopy. It requires an expert just to mark the beginning and end of the pathology in the video based on the snapshot, automatically detects bounding boxes of the pathologies (in our case polyps) within all the frames of the marked video sequences and offers a non-expert annotator comfortable fine-tuning of the bounding boxes if necessary. Figure 1 summarizes the workflow of our tool.

2. Methods

First, the user inputs a video file to our annotation tool. All the frames of the video are extracted and stored as images in the referencing folder. Then, special frames are selected by an automated domain-specific process these frames are then marked by an expert.



Figure 1. Workflow of our proposed annotation process.

¹ Corresponding Author, Adrian Krenzer, Julius-Maximilian University of Würzburg, Department of Artificial Intelligence and Knowledge Systems, Sanderring 2, 97070 Würzburg, Germany; E-mail: adrian.krenzer@uni-wuerzburg.de.

	Quality (%)		Speed (FPS)		Speed (min)		Videoinfomation		
	CVAT	Ours	CVAT	Ours	CVAT	Ours	Frames	Polyps	Framesize
Video 1	99.16	99.04	9.36	3.64	57.88	22.5	371	1	384x288
Video 2	99.44	99.58	5.31	3.81	39.74	28.5	449	1	384x288
Video 3	-	-	2.72	1.35	57.98	28.78	1279	1	898x720
Mean	99.3	99.31	5.79	2.93	51.86	26.59	700	1	555x432

Table 1. Comparison of the well-known CVAT annotation tool to our new annotation tool for faster annotation for gastroenterologists. Videos 1 and 2 are open source and annotated. Video 3 is from a German clinic. For video 3 no quality evaluation is performed since there are no ground truth annotations available. The quality metric is the mean average precision (mAP70) when the drawn box matches the ground truth box to 70 %.

The model is trained to predict polyp bounding boxes on the marked frames. This model can always be retrained with the newly annotated data. As the detection gets better, the time for annotation should decrease. Therefore, the annotation time will be reduced with annotation progress. For the detection of the polyps state of the art object detector YOLOv4 [2] is used. The detector is trained on openly available datasets for polyp detection. In the final step, a non-expert corrects the model's predictions. He gets all the model's predictions and one or more annotated boxes from the expert. Those annotations are still of high quality as the expert annotations assure the presence of a pathology and the non-experts only adjust the predicted bounding boxes.

3. Preliminary Evaluation

For our preliminary evaluation, two test subjects are instructed to use our annotation tool and the state of the art annotation tool CVAT [3]. The test subjects in this experiment are undergraduates from the field of computer science therefore just the non-expert part of our tool is evaluated. Both students are instructed to annotate the polyp frames as fast and as accurately as they can. The results are shown in Table 1. The quality evaluation results show that almost similar annotation results to those of gastroenterology experts are achieved. For speed, our tool outperforms the CVAT tool in any video. In two videos our tool is more than twice as fast as the CVAT tool. A total of three videos were annotated. We plan a full evaluation to further investigate the speed of the expert annotators.

4. Conclusion

All in all, a novel tool for machine learning video annotation in endoscopic recordings is presented. The method's annotation speed exceeds the classic computer science tool CVAT [3] while maintaining high-quality results.

References

- [1] Hamet P, Tremblay J. Artificial intelligence in medicine. Metabolism. 2017 Apr 1;69:S36-40.
- Bochkovskiy A, Wang CY, Liao HY. YOLOv4: Optimal Speed and Accuracy of Object Detection. arXiv preprint arXiv:2004.10934. 2020 Apr 23.
- [3] Sekachev B et al.: Opencv/cvat: v1.1.0. https://github.com/opencv/cvat 2020 May 25.