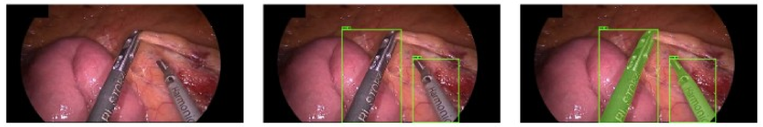
Pixel-accurate Segmentation of Surgical Tools based on Bounding Box Annotations

George Leifman[1], Amit Aides[2], Tomer Golany[1], Daniel Freedman[1], Ehud Rivlin[1], Petachia Reissman[3]
[1] Verily Life Sciences, [2] Google Research, [3] Shaare-Zedek Medical Center

verily

Objective

Detection and segmentation of surgical instruments in laparoscopic surgeries is a useful intermediate task for the development of computer-assisted surgery systems, such as pose estimation, surgical phase estimation, enhanced image fusion, video retrieval and others. These algorithms form the core for surgical assistance systems that range from improving the surgery's outcome, through operating room optimization, to surgeon training. We propose a deep learning-based approach to instrument segmentation, which addresses the binary segmentation problem in which every pixel in an image is labeled as instrument or background. The key novelty of our approach relates to the use of training data which is inexpensive and fast to acquire.

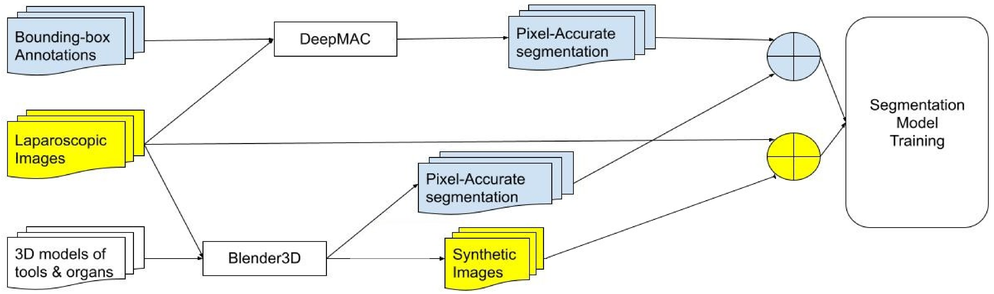


(a) (b) (c)

Our approach relies on weak and fast annotations provided as bounding boxes of the instruments. Given an image (a), we annotate a bounding box around each instrument (b). Then we automatically generate segmentation for each instrument based on the bounding box (c).

Methods

Given a dataset of laparoscopic images and their corresponding bounding box annotations, we generate pixel-accurate segmentations using DeepMAC. Synthetic images are generated by Blender3D. To train the segmentation model, the above types of data are combined. Images are in yellow, annotation and automatic masks are in blue. The crossed circles combine the read data with the synthetic one.



Results

We compare our results to SOTA on EndoVis2017 and EndoVis2019 datasets. When no pixel-accurate segmentation annotations is provided, we automatically generate the segmentation for training data. The accuracy of this model is similar to SOTA. Furthermore, incorporating our synthetic data into training significantly improves the results, achieving SOTA results based on weak annotations only.

SEGMENTATION RESULTS ON ENDOVIS 2017 DATASET

Training			Testing
Method	Annotation	Synthetic Data	(Dice)
Shvets et al. 2018 [23]	pixel-accurate	-	0.900
RASNet 2019 [43]	pixel-accurate	-	0.946
Our	pixel-accurate	-	0.931
Our	pixel-accurate	5K	0.943
Our	-	5K	0.861
Our	bounding box	-	0.898
Our	bounding box	5K	0.939

SEGMENTATION RESULTS ON ENDOVIS 2019 DATASET FOR 3 TESTING STAGES.

Training			Testing (Dice)		
Method	Annotation	Synth. Data	Stage 1	Stage 2	Stage 3
Or-unet 2020 [16]	pixel-accurate	-	0.92	0.90	0.88
Our	pixel-accurate	-	0.90	0.88	0.87
Our	pixel-accurate	5K	0.91	0.89	0.89
Our	-	5K	0.71	0.74	0.72
Our	bounding box	-	0.88	0.89	0.85
Our	bounding box	5K	0.90	0.88	0.89

Conclusions

We introduced a novel approach for instrument segmentation in laparoscopic surgeries that relies on weak and fast annotations provided as bounding boxes of the surgical instruments. To achieve an accurate automatic generation of the segmentation based on bounding boxes only, we employed the recently proposed DeepMAC method. Our approach achieved SOTA results based on weak annotations only. The evaluation was performed on various instrument segmentation datasets proposed during the last few years. In the future we plan to extend our work from binary segmentation to instrument and instrument part segmentation. We also plan to explore how our synthetic data can enable the 3D pose estimation of the instruments.