Self-supervised Artificial Intelligence for Laparoscopic Surgeries

Roy Hirsch, Mathilde Caron, Regev Cohen, Amir Livne, Tomer Golany,
Daniel Freedman and Ehud Rivlin.

verily

Background

Current Machine Learning (ML) algorithms for computer-assisted surgery require large amounts of task-specific annotations which are prohibitively expensive and time-consuming.

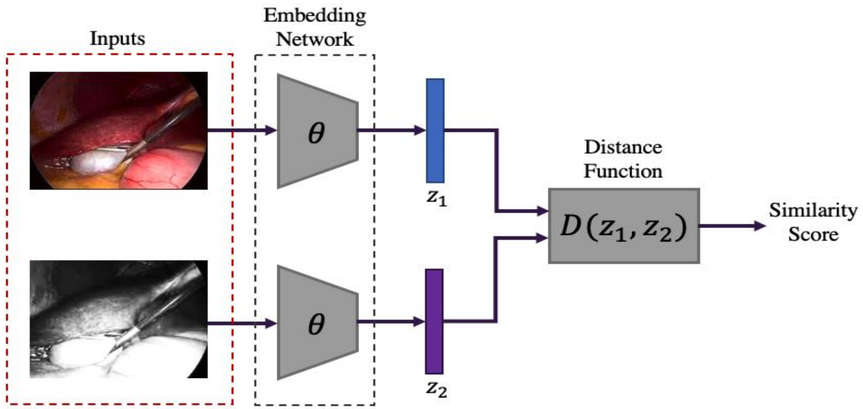
We introduce an ML framework for surgical phase recognition that relies on self-supervised learning (SSL), a type of unsupervised learning which significantly reduces the need for annotated surgical data.

Dataset

We focus on Laparoscopic Cholecystectomy (LC) procedures.

We gathers an *unannotated proprietary* dataset of 1353 LC videos of various complexity from 5 hospitals for the SSL training. We evaluated the model performances over the *public* Cholec80 dataset [1].

Figure 1: MSN model's architecture



Method

The phase-recognition system is trained in two stages:

1) **SSL Pre-Training:** We train on our *proprietary dataset* with no annotations using a self-supervised method called DINO [2] which performs knowledge distillation by enforcing similarity of randomly augmented set of views of the same image (see Figure 1).

2) **Supervised Fine-Tuning:** We use the informative image representations from the pre-trained models for training a linear classifier over Cholec80 dataset.

Results

We evaluate our ML system on 32 Cholec80 test videos, and report the per-frame accuracy and macro F1 of our model DINO-X where X indicates the number of annotated Cholec80 videos used for fine-tuning.

We compare our method performance to:
1) *Fully-supervised* network trained over all of Cholec80 train fold.
2) Off-the-shelf model *fine-tuned* over all of Cholec80 train fold.

Figures 2 shows that with **12.5%-25%** of the annotations we outperform the performance of standard fine-tuning, highlighting the effectiveness of the SSL pre-training.

Figure 1a: Mean accuracy results

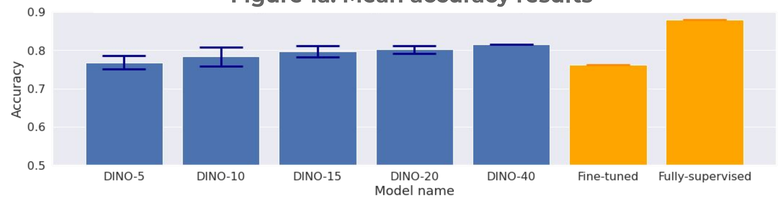
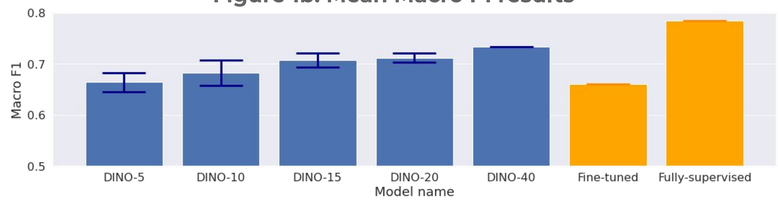


Figure 1b: Mean Macro F1 results



Conclusions

We developed an ML framework for laparoscopic surgeries which leverages unannotated datasets, learning from dramatically fewer labels. The performance of our AI phase-recognition model demonstrates the data-efficiency of our framework, significantly impacting the annotation and learning processes.

[1] Twinanda, Andru P., et al. "Endonet: a deep architecture for recognition tasks on laparoscopic videos." IEEE transactions on medical imaging 36.1 (2016): 86-97.

[2] Assran, Mahmoud, et al. "Masked siamese networks for label-efficient learning." Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXXI. Cham: Springer Nature Switzerland, 2022.