Surgical Pre-training for Activity Recognition to learn efficiently from Limited Data

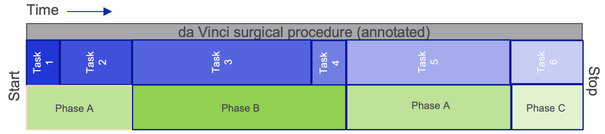
Sreeram Kamabattula,
Kiran Bhattacharyya, and
Anthony Jarc

Motivation

Automated detection of surgical activities in procedure videos can enable surgeons to review and analyze their cases efficiently. However, translating these activity recognition models into clinical practice becomes challenging as their performance is limited by availability of clean annotated datasets.

Introduction

Surgeries can be studied at different granular levels; they can be deconstructed into clinically-relevant surgical activities performed during a surgical procedure such as phases or tasks.



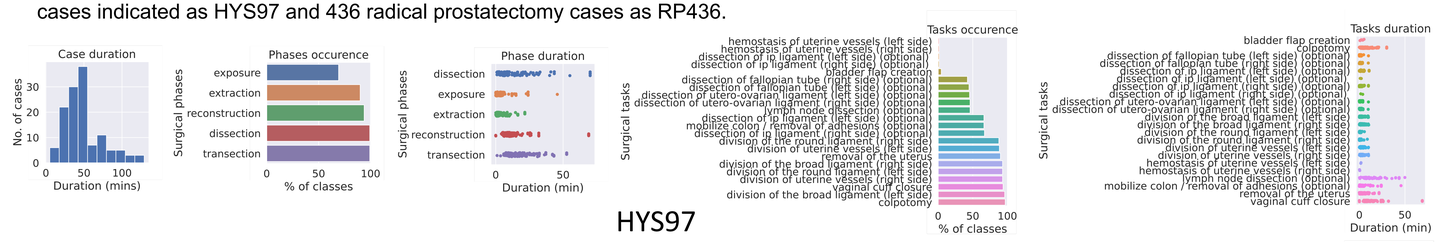
Surgical tasks are more granular than Surgical phases

However, manual annotation of these surgical activities in procedure videos becomes difficult and takes longer with increasing granularity. Thus, while a lot of surgical data (procedure videos) is collected, the models are trained using only the portion of data that is annotated with concerned activities. This little data could result in poor detection of surgical activities.

We believe that leveraging all the surgical data for pre-training could result in better prediction of surgical activities as some salient features can be learned. We present a scenario where pre-training on surgical phases can be beneficial for predicting surgical tasks.

Datasets

We demonstrate the performance on two robot-assisted surgical datasets of different sizes and statistics as shown in below plots. 97 hysterectomy cases indicated as HYS97 and 436 radical prostatectomy cases as RP436.



HYS97



RP436

Phases are agnostic to surgical procedures where tasks are procedure-specific. Each phase or task occurrence can be optional in a surgical procedure and can range from seconds to minutes.

Method

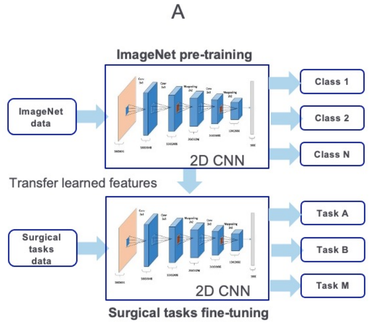
Fig A) Illustration of pre-training on a non-surgical ImageNet [1] dataset and fine-tuning to surgical tasks.

ImageNet is widely adopted for pre-training and is our baseline.

Fig B) Surgical pre-training --- ImageNet pre-trained model is first fine-tuned to learn surgical context using phase-annotated data. Then, this phase pre-trained model is again fine-tuned to learn surgical tasks.

Both HYS97 and RP436 datasets are phase-annotated, and first individual phase pre-trained models are developed as described above.

To evaluate advantages of ImageNet and surgical phase pre-training, we consider various cases where only a certain ratio of the datasets are annotated with tasks for fine-tuning.



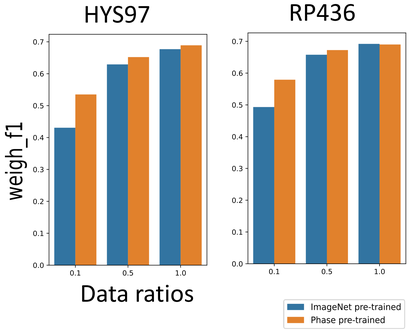
Note that these are 2DCNN image-based models. Temporal information is not used

Results

Average performance of detecting surgical tasks with surgical phase (orange) or ImageNet (blue) pre-training is shown for three data ratios on the right figure.

Data ratio refers to amount of data annotated with tasks. 0.1 refers to an assumption that only 10% of the data is annotated with tasks, 0.5 is 50%, and 1.0 is 100%.

Surgical phase pre-training demonstrates improved recognition of surgical tasks upon baseline. We believe that the performance gain increases a) as pre-training data increases, b) due to high correlation between surgical phases and tasks.



Conclusion

Our goal is to provide a proof of concept for benefits of surgical pre-training for surgical activity recognition. Above results suggest that surgical pre-training can be significant when leveraging a lot of data for pre-training. This is also an example showing that pre-training and learning surgical context from other surgical contexts could be beneficial

[1] Deng, J. et al., 2009. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*. pp. 248–255.