Learning from a Tiny Dataset of Manual Annotations: a Teacher/Student Approach For Surgical Phase Recognition

Tong Yu[1], Didier Mutter[2], Jacques Marescaux[2], and Nicolas Padoy[1]
[1] ICube, University of Strasbourg, CNRS, IHU Strasbourg, France
[2]University Hospital of Strasbourg, IRCAD and IHU Strasbourg, France

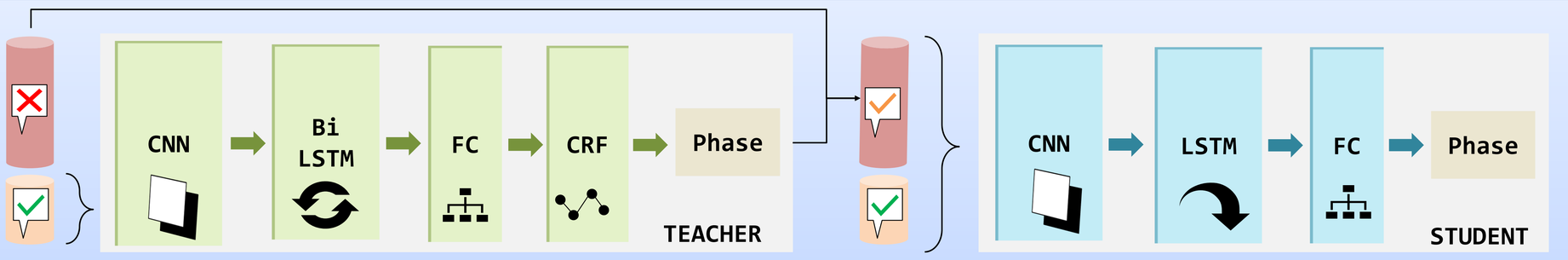
Abstract

Vision algorithms capable of interpreting scenes from a real-time video stream are necessary for computer-assisted surgery systems to achieve context-aware behavior. In laparoscopic procedures one particular algorithm needed for such systems is the identification of surgical phases, for which the current state of the art is a model based on a CNN-LSTM. A number of previous works using models of this kind have trained them in a fully supervised manner, requiring a fully annotated dataset. Instead, our work confronts the problem of learning surgical phase recognition in scenarios presenting scarce amounts of annotated data (under 25% of all available video recordings). We propose a teacher/student type of approach, where a strong predictor called the teacher, trained beforehand on a small dataset of ground truth-annotated videos, generates synthetic annotations for a larger dataset, which another model - the student - learns from. In our case, the teacher features a novel CNN-biLSTM-CRF architecture, designed for offline inference only. The student, on the other hand, is a CNN-LSTM capable of making real-time predictions. Results for various amounts of manually annotated videos demonstrate the superiority of the new CNN-biLSTM-CRF predictor as well as improved performance from the CNN-LSTM trained using synthetic labels generated for unannotated videos.

Overview

- TASK: identify the surgical phase

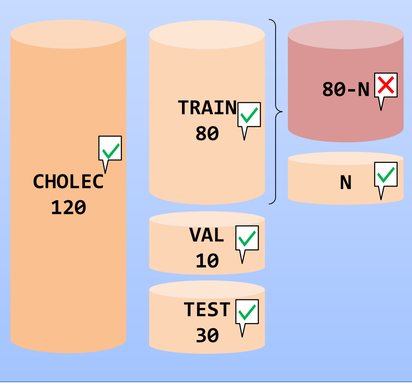
#	Name
1	Preparation
2	Calot triangle dissection
3	Clipping and cutting
4	Gallbladder dissection
5	Gallbladder packaging
6	Cleaning and coagulation
7	Gallbladder extraction

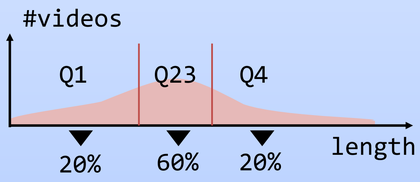


- Perform **phase recognition** under **extreme scarcity** scenarios
  - annotations provided for **≤ 25 %** of available videos
- Introduce a new **offline teacher** predictor that performs reasonably well under those circumstances
  - **CNN - biLSTM – CRF** architecture
- Use it to generate **synthetic phase labels** for training a **real-time student** predictor
  - **CNN – LSTM** architecture

Dataset: cholec120

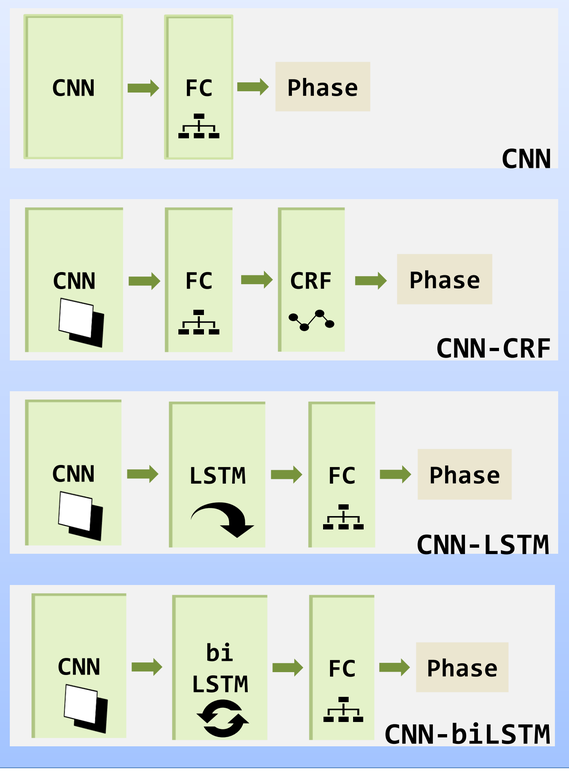
- 120 full-length videos of laparoscopic cholecystectomy
- Average runtime / video: 2288 ± **960** seconds
- Number of available annotated videos for training limited to **mini-training sets** of size **N**
- Random mini-training set sampling based on video length
- **3** distinct mini training sets for every size N



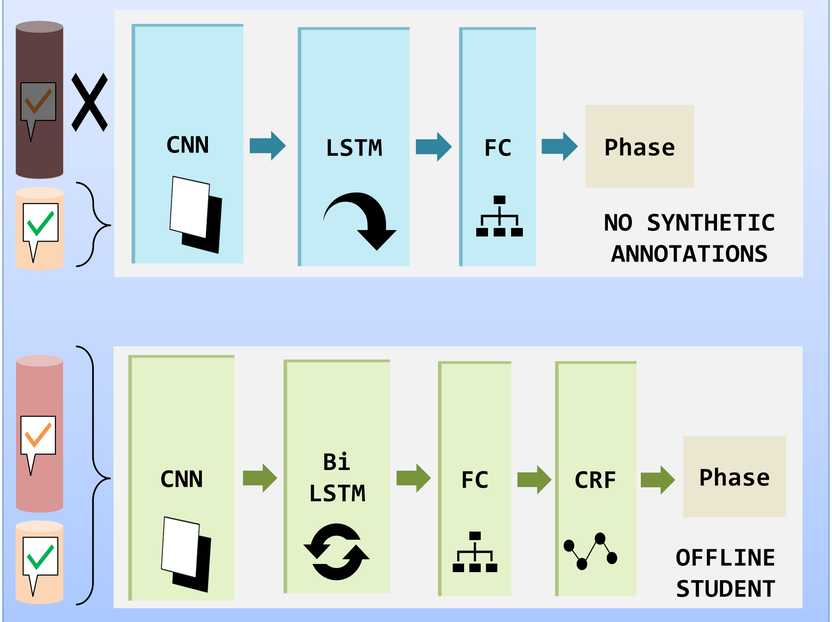


# annotated videos (N)	Q1	Q23	Q4
1	0	1	0
3	1	1	1
5	1	3	1
10	2	6	2
20	4	12	4

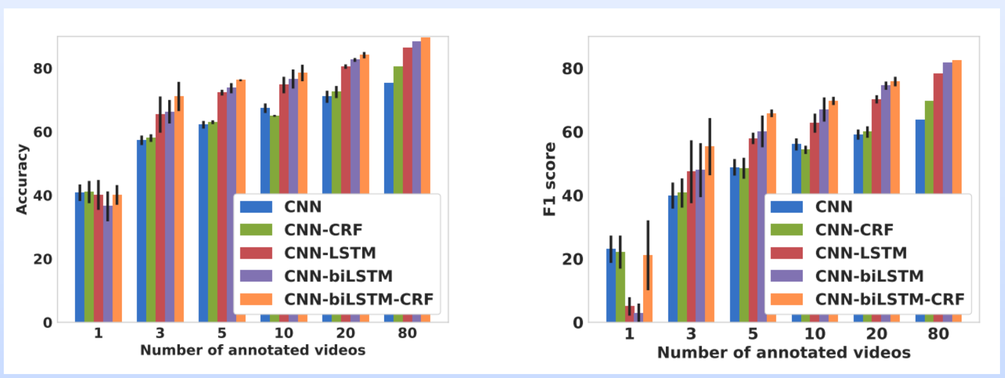
Teacher ablation studies



Student experiments



Teacher results



Student results

# of ground-truth annotated videos used		3	5	10	20
No synthetic annotations	Acc	65.4 ± 5.7	72.3 ± 0.9	74.7 ± 2.6	80.5 ± 0.7
	F1	47.4 ± 9.9	57.9 ± 1.8	62.7 ± 3	70.2 ± 1.3
Student	Acc	74.6 ± 3.6	77.7 ± 0.8	79.1 ± 0.9	83.4 ± 0.3
	F1	56.1 ± 8.6	64.5 ± 2.9	66.9 ± 3.1	73.2 ± 0.8
Teacher	Acc	71.1 ± 4.6	76.2 ± 0.3	78.5 ± 2.6	84.1 ± 1
	F1	55.3 ± 9	65.8 ± 1.2	69.7 ± 1.3	75.8 ± 1.5
Offline student	Acc	74.9 ± 4.6	78.7 ± 1	80.8 ± 0.8	86.3 ± 1
	F1	55.7 ± 9.7	66.8 ± 2.7	69.9 ± 4	78.1 ± 1

Conclusion

- Superiority of the CNN – biLSTM –CRF teacher model for offline inference
- Improvement from online models trained with synthetic annotations generated by the teacher
- Potential for scaling to more cholecystectomy videos
- Potential for adapting to more complex surgery types

References

- Charrière, K., Quellec, G., Lamard, M., Martiano, D., Cazuguel, G., Coatrieux, G., Cochener, B.: Real-time analysis of cataract surgery videos using statistical models. Multimedia Tools Appl. (2017)
- Funke, I., Jenke, A., Mees, S.T., Weitz, J., Speidel, S., Bodenstedt, S.: Temporal coherence-based self-supervised learning for laparoscopic workflow analysis. In: OR 2.0 Context-Aware Operating Theaters, Computer Assisted Robotic Endoscopy, Clinical Image-Based Procedures (2018)
- Huang, Z., Xu, W., Yu, K.: Bidirectional LSTM-CRF models for sequence tagging (2015)
- Jin, Y., Dou, Q., Chen, H., Yu, L., Qin, J., Fu, C.W., Heng, P.A.: Sv-rcnet: Workflow recognition from surgical videos using recurrent convolutional network. IEEE transactions on medical imaging (2018)
- Yengera, G., Mutter, D., Marescaux, J., Padoy, N.: Less is more: Surgical phase recognition with less annotations through self-supervised pre-training of cnn-lstm networks. (2018)