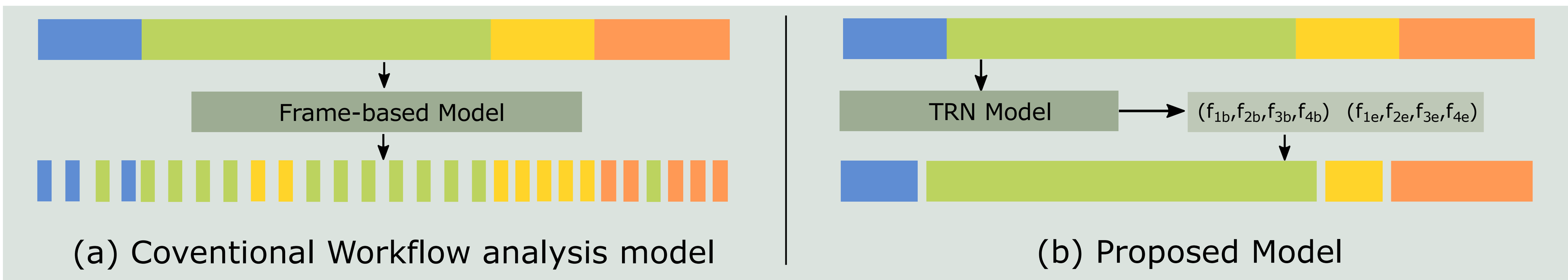
Retrieval of Surgical Phase Transitions using Reinforcement Learning

Yitong Zhang[1], Sophia Bano[1], Ann-Sophie Page[2], Jan Deprest[2], Danail Stoyanov[1], Francisco Vasconcelos[1]

[1] Wellcome/EPSRC Centre for Interventional and Surgical Sciences(WEISS) and Department of Computer Science, University College London, London, UK; [2] Department of Development and Regeneration, University Hospital Leuven, Leuven, Belgium

1. Introduction

The conventional approach to surgical workflow segmentation is to determine the surgical phase of each individual frame in a video recording of the operation, as a multi-class classification problem.



Comparison with conventional methods

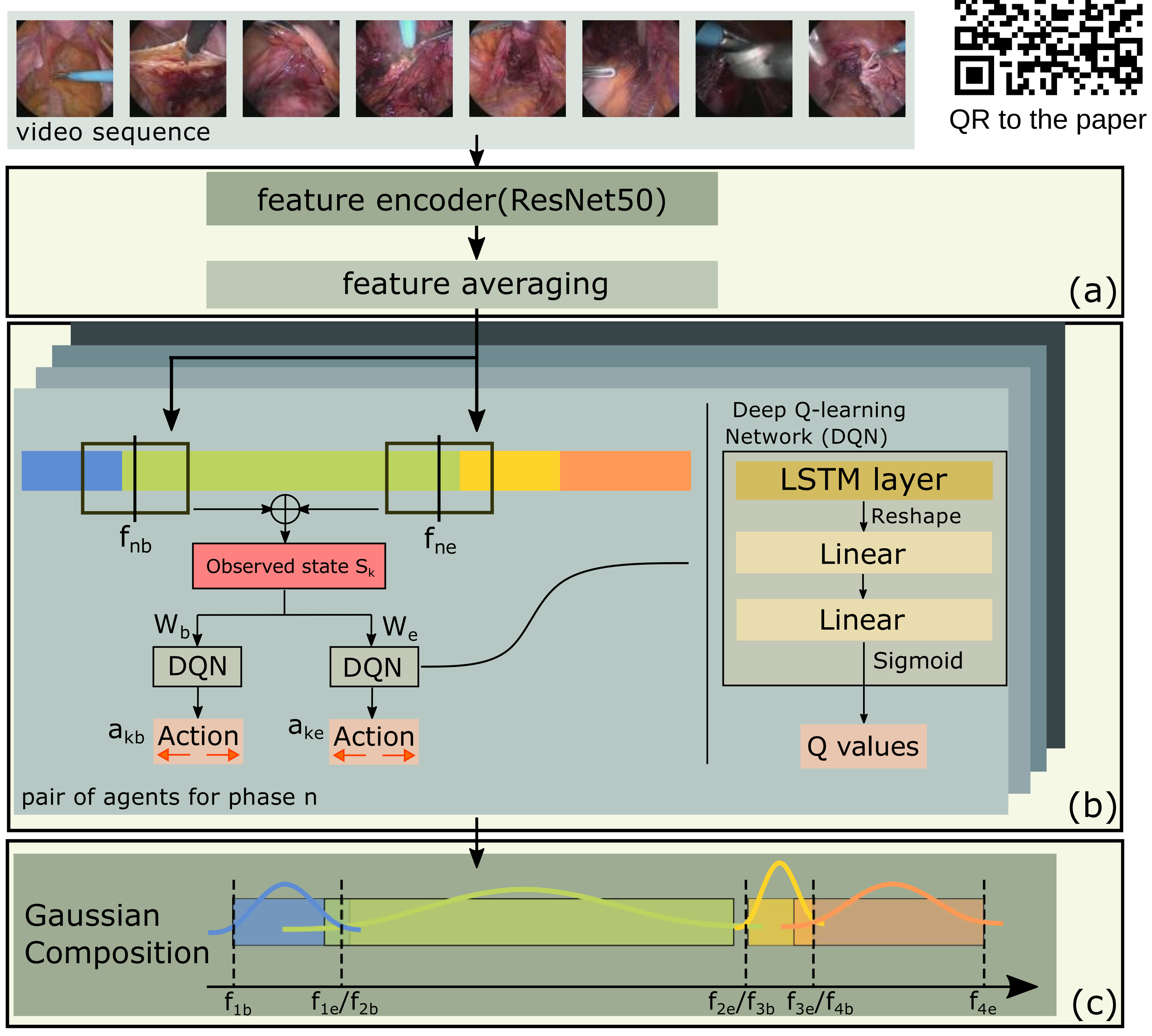
2. Contribution

We introduce a novel reinforcement learning approach for offline workflow segmentation that finds the time of phase transitions instead of classifying individual frames. We investigate two configurations with different search window sizes (21, 41 and 81 clips) for two datasets:

- Fixed initialization (FI): needs only a portion of video (<60% and <20% for 2 datasets), saving significant computational cost
- ResNet Modified Initialization (RMI): outperforms the state-of-the art (TeCNO[2] and Trans-SV[1]) at a comparable computational cost.

3. Proposed Method

QR to the paper



TRN architecture with (a) averaged ResNet feature extractor, (b) multi-agent network for transition retrieval and (c) Gaussian composition operator

4. Results

Datasets:

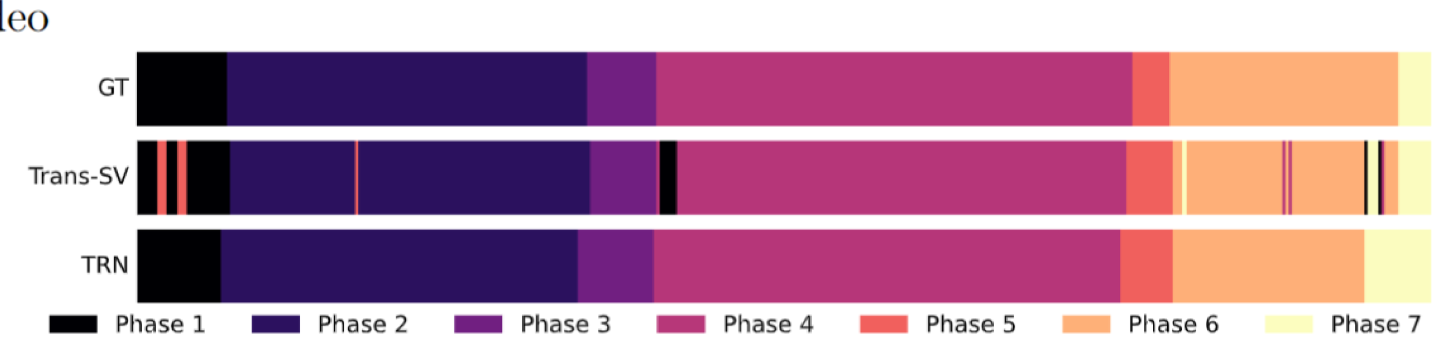
- Cholec80 [3]: 7 phases, average duration of 41 mins
- Sacrocolpopexy [4] divided into suturing/non-suturing phases as a clinical interest, average duration of 3 hours and 13 mins

Window size	Phase 1	Phase 2	Phase 3	Phase 4	Phase 5	Phase 6	Phase 7	Overall F1-score
TRN21 FI	0.854	0.917	0.513	0.903	0.687	0.549	0.83	0.782
TRN41 FI	0.828	0.943	0.636	0.922	0.558	0.694	0.85	0.808
TRN21 RMI	0.852	0.942	0.619	0.939	0.727	0.747	0.837	0.830
TRN41 RMI	0.828	0.940	0.678	0.945	0.753	0.738	0.861	0.846

Table 1: TRN ablation in the Cholec80 dataset (F1-scores). The values per-phase are computed before Gaussian Composition, while the overall F1-score is for the complete TRN method.

Dataset	Method	Accuracy	F1-Score	Event ratio	Ward Event Ratio	Coverage rate(%)	Computatio--nal Cost(s)
Cholec80	ResNet-50	79.7±7.5	0.756	0.120	0.375	full	96.6
	TeCNO	88.3±6.5	0.774	0.381	0.691	full	99.6
	Trans-SVNet	89.1±5.7	0.800	0.316	0.566	full	99.6
	TRN21 FI	85.3±9.6	0.782	1	0.934	57.6	60.6
	TRN41 FI	87.8±8.1	0.808	1	0.956	59.1	64.9
	TRN41 RMI	90.1±5.7	0.846	1	0.985	full	105.5
Sacrocol--popexy	ResNet-50	92.5±3.8	0.892	0.029	0.016	full	493.7
	TeCNO	98.1±1.7	0.976	0.136	0.438	full	493.8
	Trans-SVNet	97.8±2.2	0.971	0.536	0.813	full	493.9
	TRN21 FI	89.8±6.2	0.860	0.971	0.875	14.6	78.1
	TRN81 FI	90.7±6.1	0.875	0.941	0.860	18.3	104.0

Table 2: Evaluation metric results summary of ResNet-50, our implementation of TeCNO and Trans-SV, and ablative selected TRN result on Cholec80 and Sacrocolpopexy. The computatinla cost is in average second to process a single video



(a) An example of video77 from Cholec80 processed by Trans-SV and TRN41 RMI

(b) An example video from Sacrocolpopexy processed by Trans-SV and TRN81 FI

5. Conclusion

The FI configuration yields comparable results with low computational cost, while the RMI configuration yields best results.

Advantages:

- avoiding frame-level noise in predictions and enforcing phases to be continuous blocks.
- partial coverage (FI) reducing the computational cost

Limitations:

- phases with unknown number of repetitions
- not suitable for real-time operation at this moment
- Scalability issues with large number of phases

Future work will adapt more advanced RL algorithms to overcome these limitations (e. g. actor-critic, multi-agent)

References

[1] Gao, Xiaojie, et al. "Trans-SVNet: accurate phase recognition from surgical videos via hybrid embedding aggregation transformer." MICCAI, 2021..

[2] Czempiel, Tobias, et al. "TECNO: Surgical phase recognition with multi-stage temporal convolutional networks." MICCAI, 2020.

[3] 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.

[4] Claerhout, Filip, et al. "Implementation of laparoscopic sacrocolpopexy—a single centre's experience." International urogynecology journal 20.9 (2009): 1119-1125.

yitong.zhang.19@ucl.ac.uk

@WEISS_UCL

Acknowledgements This research was supported by the Wellcome/EPSRC Centre for Interventional and Surgical Sciences (WEISS) [203145/Z/16/Z]; the Engineering and Physical Sciences Research Council (EPSRC) [EP/P027938/1, EP/R004080/1, EP/P012841/1]; the Royal Academy of Engineering Chair in Emerging Technologies Scheme, and Horizon 2020 FET Open (863146).