DR.SIMON : Domain-wise Rewrite for Segment Informed Medical Oversight Network

Seohyun Lee[1], Suhyun Choe[2*], Jaeha Choi[3*], Jin Won Lee[4*]

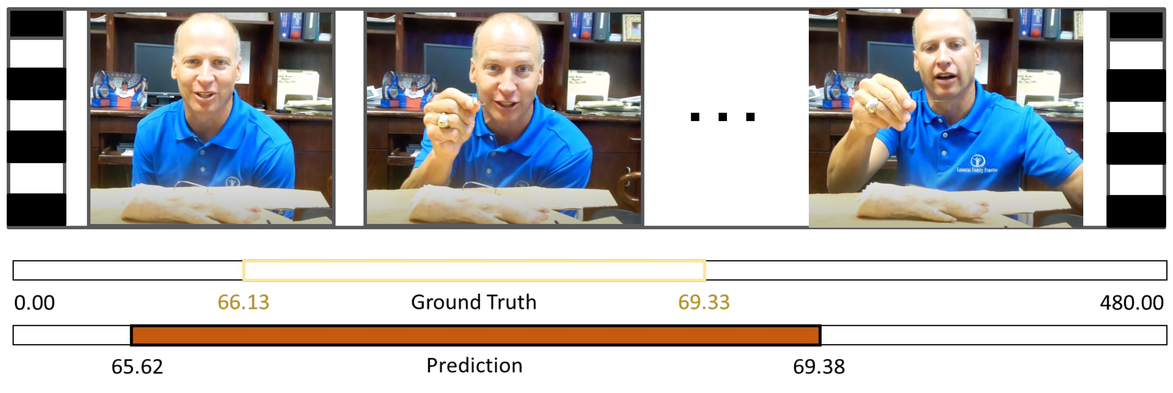
[1] Korea University, [2] Yonsei University, [3] Incheon National University, [4] McGill University

Introduction

Video Temporal Grounding in Medical Videos

: locate the start–end segment for a text query

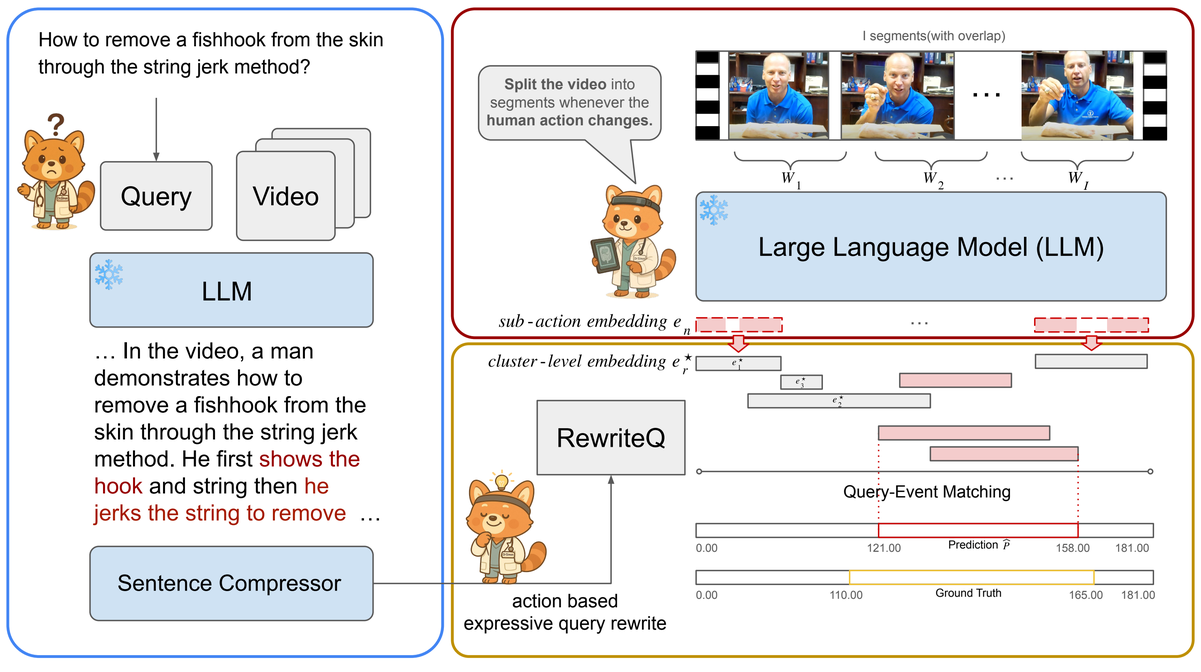
Question : How to remove a fishhook from the skin through the string jerk method?



Challenges

- Limitations in long-form video handling
- Medical queries are not clearly visible in the frame
- Requires medical-domain fine-tuning to effectively process medical queries

Pipeline Overview



- **Query Rewriting Module**

  Inject visually explicit cues into medical queries
- **Boundary Event Segmentation Module**

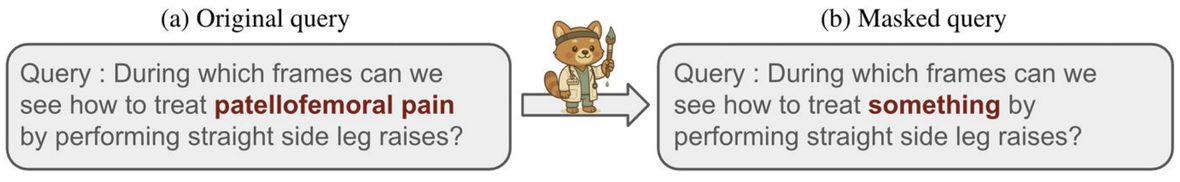
  Efficiently handle long videos
- **Query Event Matching Module**

  Reformulate grounding as action–event alignment

Motivation

Diagnostic Masking Experiment

: replaced all medical queries with “something”



Method	mIoU↑	R@0.3↑	R@0.5↑	R@0.7↑
VTimeLLM	6.32	9.86	4.93	2.11
+ Masked Query	7.92	11.97	6.34	3.52

Key Insights

- Performance did not drop when medical terms were masked
- Bottleneck lies more in the lexical gap than in visual perception
- Aligning medical terminology with visual cues can improve grounding

Contributions

- Identified the main bottleneck as the lexical gap between medical terminology and visual cues
- Demonstrated performance gains via query rewriting
- Proposed a framework that adapts to emerging medical terminology efficiently

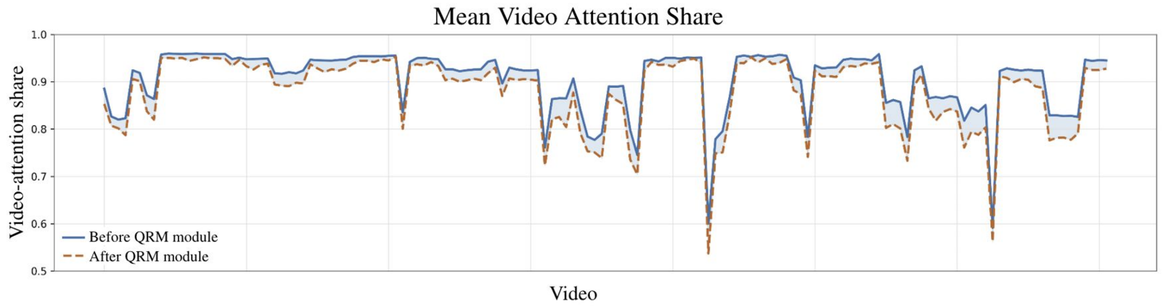
Experiment

Evaluation on MedVidCL

Method	mIoU↑	R@0.3↑	R@0.5↑	R@0.7↑
VTimeLLM	6.32	9.86	4.93	2.11
+ Masked Query	7.92	11.97	6.34	3.52
+ Rewrite Query	7.15	10.56	5.63	3.52
+ Rewrite Query (Summ.)	9.18	15.49	7.04	1.41
VSL-QGH	20.12	25.81	14.20	6.45
RevisionLLM	21.18	28.50	**26.10**	**14.28**
DR.SIMON (ours)	**28.08**	**40.14**	20.42	10.56

Verification of QRM

Setting	mIoU↑	R@0.3↑	R@0.5↑	R@0.7↑
Base FT + Orig. Q (Row 1)	14.58	21.13	9.86	4.23
Base FT + Rewrite Q (Row 2)	14.79	21.13	10.56	2.82
RewriteQ FT + Orig. Q (Row 3)	14.99	24.65	9.86	3.52
RewriteQ FT + Rewrite Q (Row 4)	**15.18**	**24.94**	**10.15**	**4.52**



Hyperparameter Sensitivity

Varying τ at $k=7$

τ	mIoU↑	R@0.3↑	R@0.5↑	R@0.7↑
1.0	22.41	32.39	12.68	7.04
0.99	**28.08**	**40.14**	**20.42**	**10.56**
0.90	25.31	30.99	15.49	9.15
0.80	22.73	24.65	10.56	4.93
0.70	21.53	21.13	9.15	4.23
0.60	19.65	16.20	9.86	4.93
0.50	18.17	14.08	7.75	3.52

Varying k at $\tau \in \{0.99, 0.90\}$

τ	k	mIoU↑	R@0.3↑	R@0.5↑	R@0.7↑
0.99	10	25.99	30.28	15.49	7.75
	7	**28.08**	**40.14**	**20.42**	**10.56**
	5	23.20	30.28	16.20	10.56
	3	17.42	23.24	12.00	5.63
	1	7.65	8.45	3.52	0.70
0.90	10	23.94	25.35	13.38	7.75
	7	25.31	30.99	15.49	9.15
	5	24.86	**33.80**	**17.61**	**9.86**
	3	20.51	26.06	13.38	7.75
	1	8.66	9.86	4.23	0.70