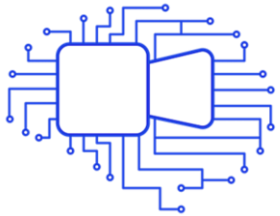


BMVC 2024

UNIVERSITY OF SURREY

UNIVERSITY OF WOLLONGONG AUSTRALIA

Interpretable Long-term Action Quality Assessment

Xu Dong[1], Xinran Liu[1], Wanqing Li[2], Anthony Adeyemi-Ejeye[1], Andrew Gilbert[1]

[1]University of Surrey, Guildford, UK

[2]Advanced Multimedia Research Lab, University of Wollongong, Wollongong, Australia

BMVC 2024

“How well these actions are performed?”

Score: 92.3

Score: 80.9

Score: 79.6

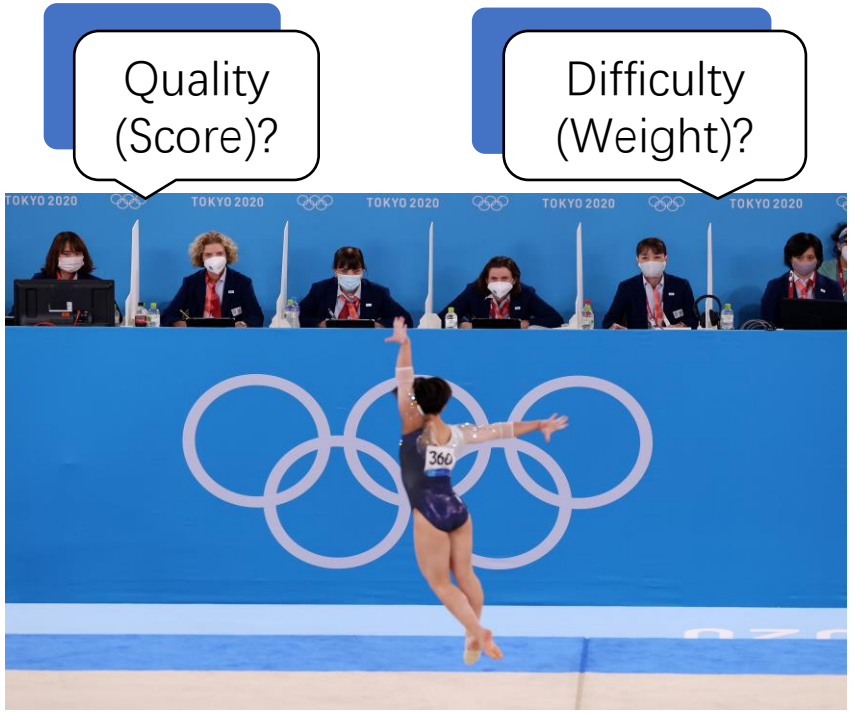
Score: 97.7

Introduction

- Real-world applications of AQA: **Sports analysis**, **Healthcare, and Daily activity assessment.**
- Eg: AQA serve as a reference for human judges in sports competitions.
- Challenges in previous work:
  - Fine-Grained Feature Extraction
  - Robustness
  - Uncertainty
  - **Interpretability**
  - **Long-term video**

Challenge One

- **Challenge:** How can we make AQA results more **interpretable**?
- Existing AQA models regress single score and lack interpretability and semantic meanings of clips.
- **Quality (Score)** of action execution and degree of **Difficulty (Weight)** are two important factors.
- How to disentangle semantic meaning of a single clip without label.

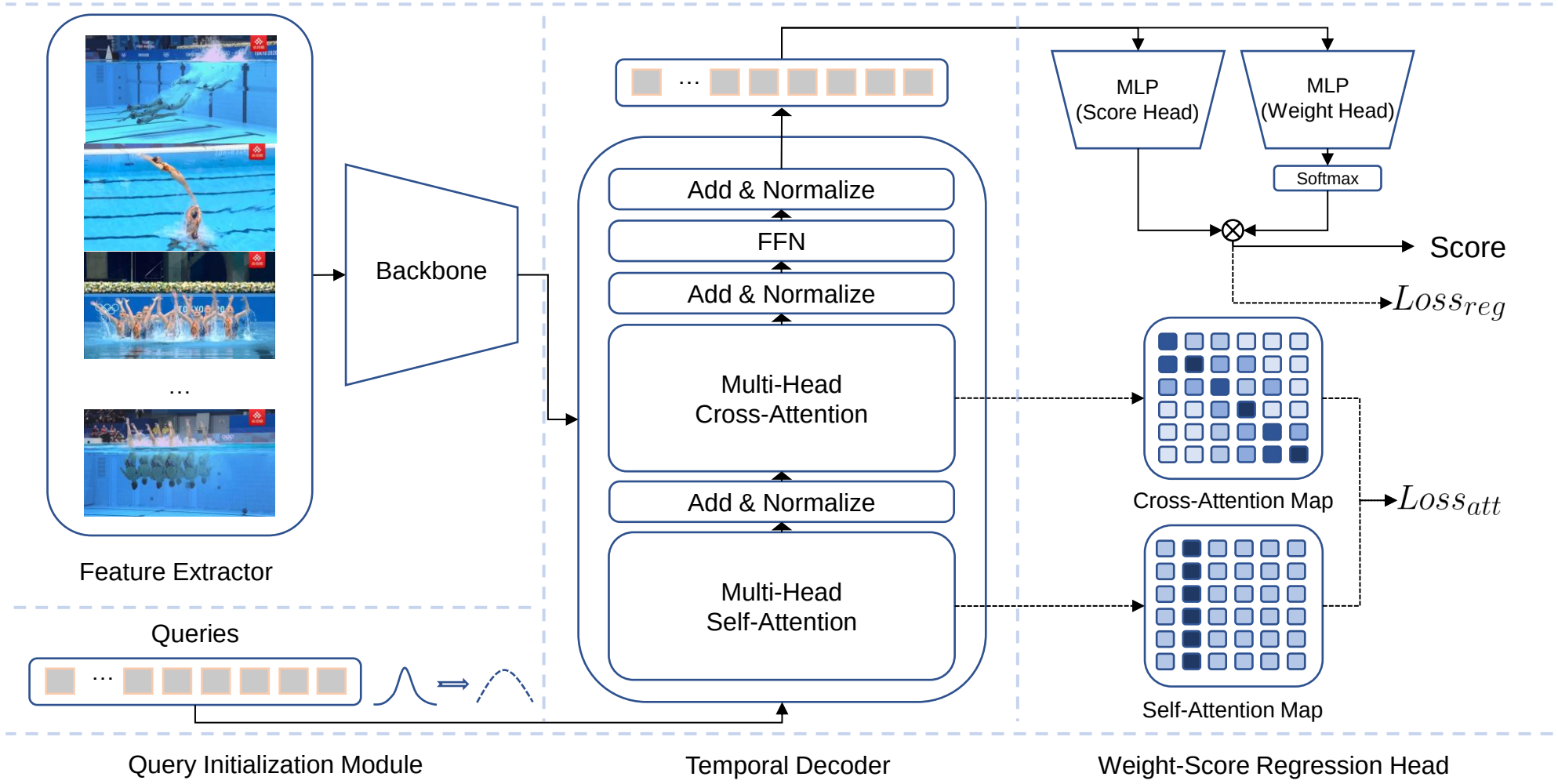


Challenge Two

- **Challenge:** Limited Understanding of Long Temporal Sequences.
- AQA datasets

Dataset	Action	Average Video Length
MTL-AQA	Diving	4.1s
FineDiving	Diving	4.2s
AQA-7-Dive	Diving	4.1s
Fis-V	**Figure Skating**	**2m 50s**
Rhythmic Gymnastics	**Gymnastics**	**1m 35s**
LOGO	**Synchronized Swimming**	**3m 24s**

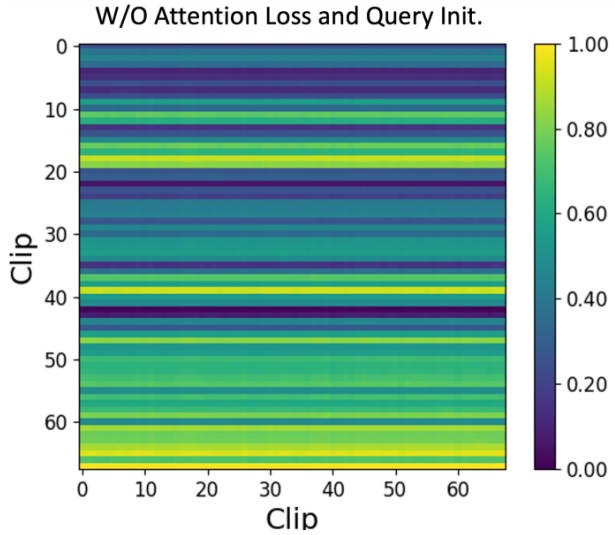
Network Structure



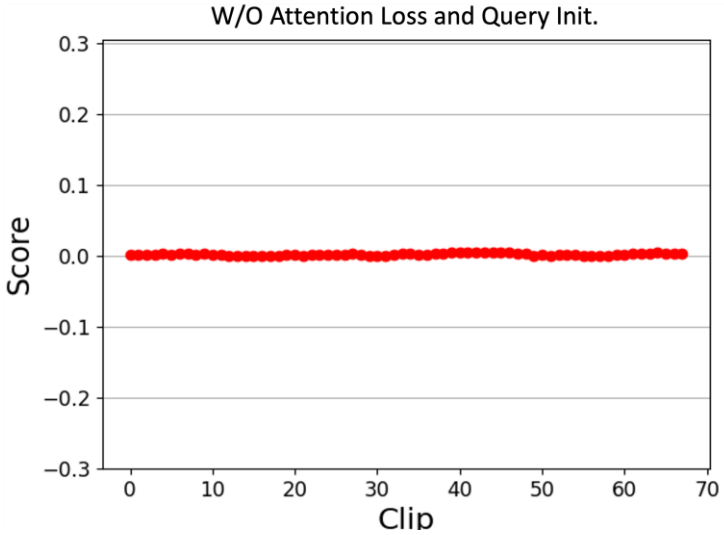
Temporal Skipping

- Temporal sequences lead the model to **select shortcuts** and skip decoder self-attention, preventing output degradation.
- Self-attention maps show near-uniform distribution.
- Each clip has averaged score in segment score.
- This can be mitigated by our proposed **Attention loss** and **Query Initialization module.**

Self-attention Map

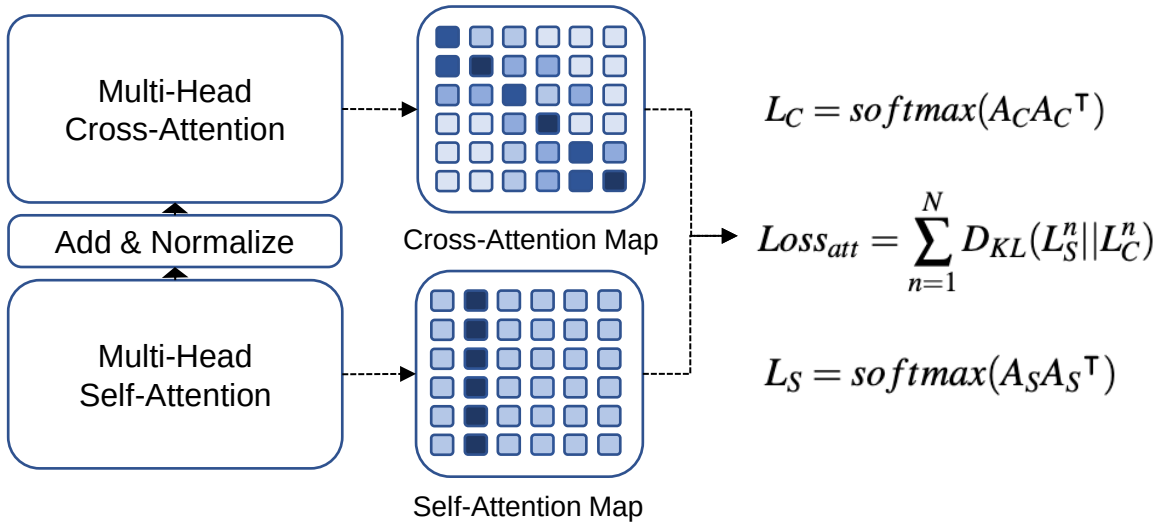


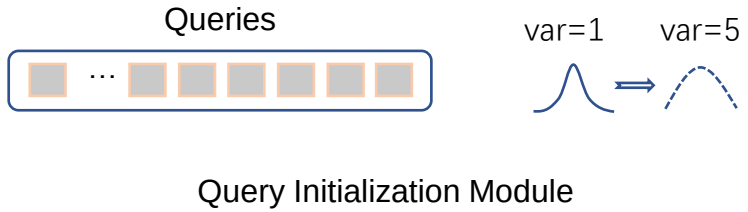
Segment Score

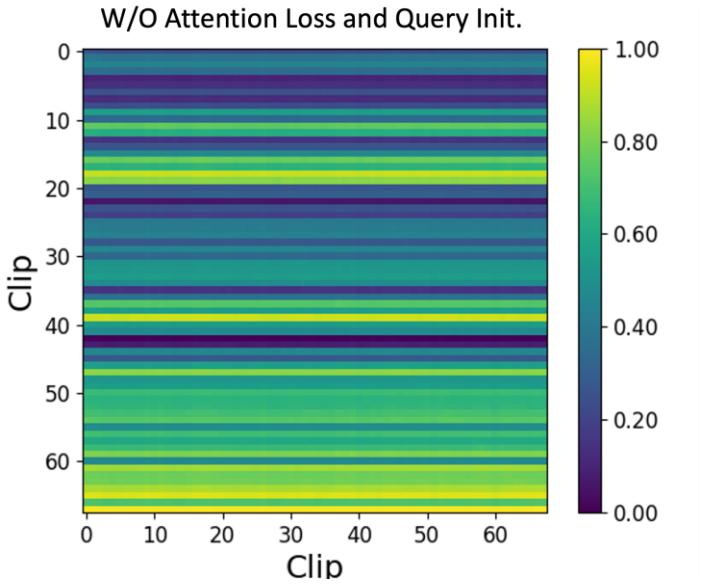


Methodology

- Attention loss uses KL Divergence to constrain Self-attention and Cross-Attention outputs.
- Query Initialization module uses larger variance to initialize query embedding in transformer decoder.

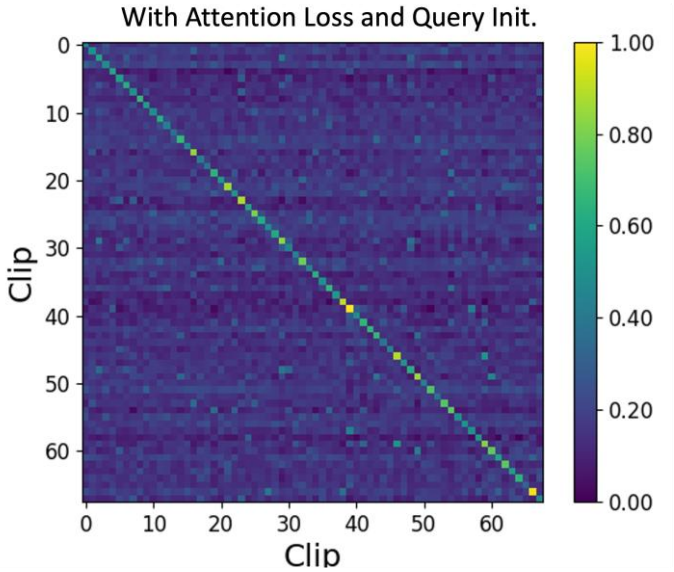


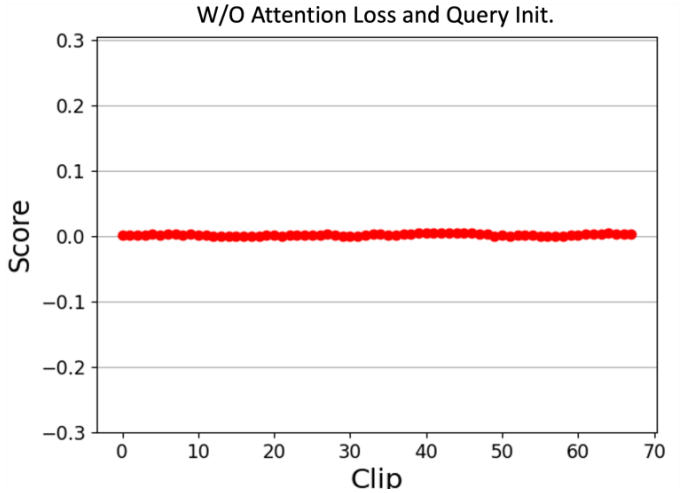


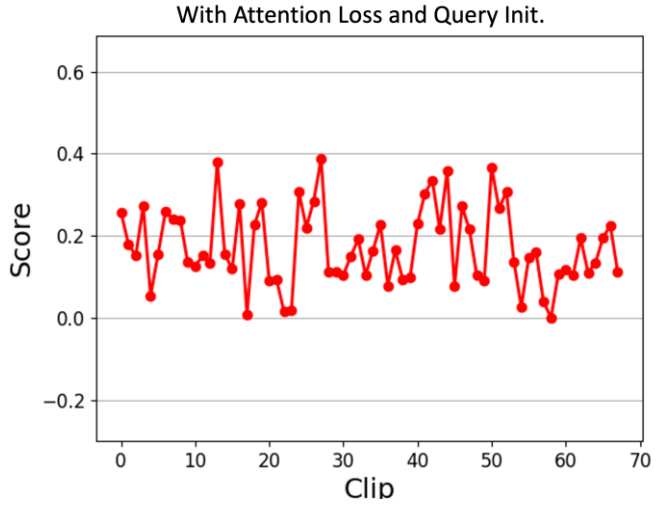


Query Init.

Attention Loss.







Experiment

Performance comparison on Rhythmic Gymnastics (RG) and Figure Skating Video (Fis-V) dataset

Methods	Feature Extractor	RG (SRCC↑)					Fis-V (SRCC↑)		
		Ball	Clubs	Hoop	Ribbon	Avg.	TES	PCS	Avg.
SVR [19]	C3D [25]	0.357	0.551	0.495	0.516	0.483	0.400	0.590	0.501
MS-LSTM [32]	I3D [3]	0.515	0.621	0.540	0.522	0.551	-	-	-
	VST [17]	0.621	0.661	0.670	0.695	0.663	0.660	0.809	0.744
ACTION-NET [35]	I3D[3]+ResNet[11]	0.528	0.652	0.708	0.578	0.623	-	-	-
	VST[17]+ResNet[11]	0.684	0.737	0.733	0.754	0.728	0.694	0.809	0.757
GDLT [31]	VST [17]	0.746	0.802	0.765	0.741	0.765	0.685	0.820	0.761
Ours	VST [17]	**0.823**	**0.852**	**0.837**	**0.857**	**0.842**	**0.717**	**0.858**	**0.788**

Performance comparison on LOGO dataset

Methods	I3D [3]		VST [17]	
	SRCC ↑	R-$\ell 2$(×100) ↓	SRCC ↑	R-$\ell 2$(×100) ↓
USDL [24]	0.426	5.736	0.473	5.076
CoRe [34]	0.471	5.402	0.500	5.960
TSA [33]	0.452	5.533	0.475	4.778
ACTION-NET [35]	0.306	5.858	0.410	5.569
USDL-GOAT [38]	0.462	4.874	0.535	5.022
TSA-GOAT [38]	0.486	5.394	0.484	5.409
CoRe-GOAT [38]	0.494	5.072	0.560	4.763
Ours	**0.593**	**1.220**	**0.780**	**1.745**

Ablation Study

Ablation study on the average performance across various modules.

Module	Attention Loss	Query PE	Query Init.	SRCC ↑
Baseline	×	×	×	0.628
	✓	×	×	0.807
	✓	✓	×	0.810
Ours	✓	✓	✓	**0.842**

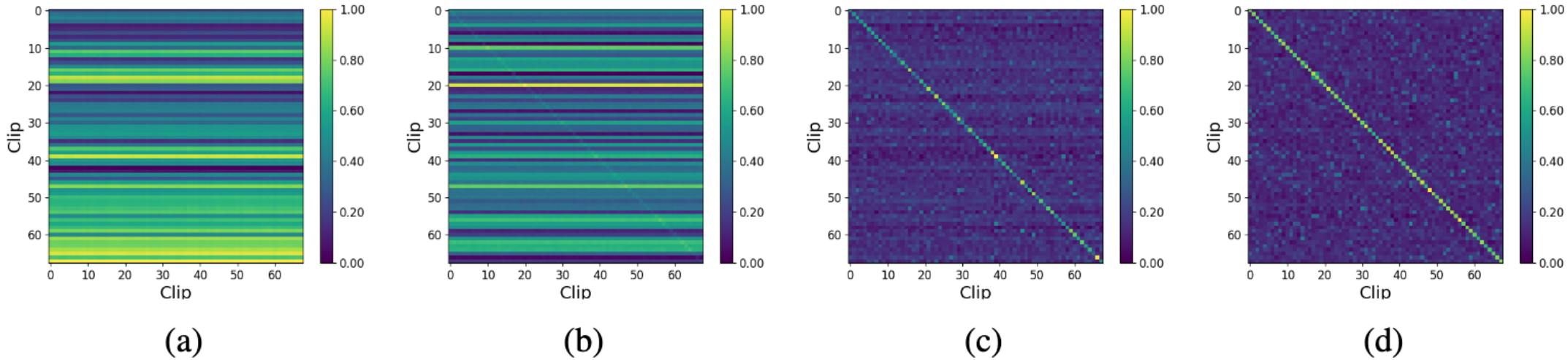
Effect of Positional Encoding on RG dataset

Methods	Query	Memory	SRCC
Baseline	×	×	0.758
	×	✓	0.778
	✓	✓	0.751
Ours	✓	×	**0.824**

Effect of Query Variance Initialization on RG dataset

Variance Init.	SRCC
0.5	0.810
1	0.810
3	0.811
5	**0.820**

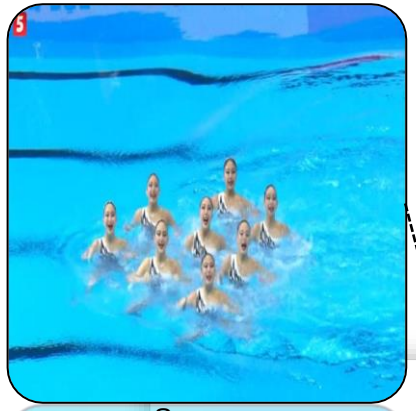
Query Initialization



Self-attention map of query initialized with different variances

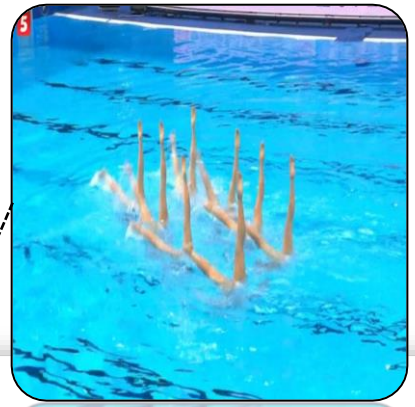
Visualization of Interpretability

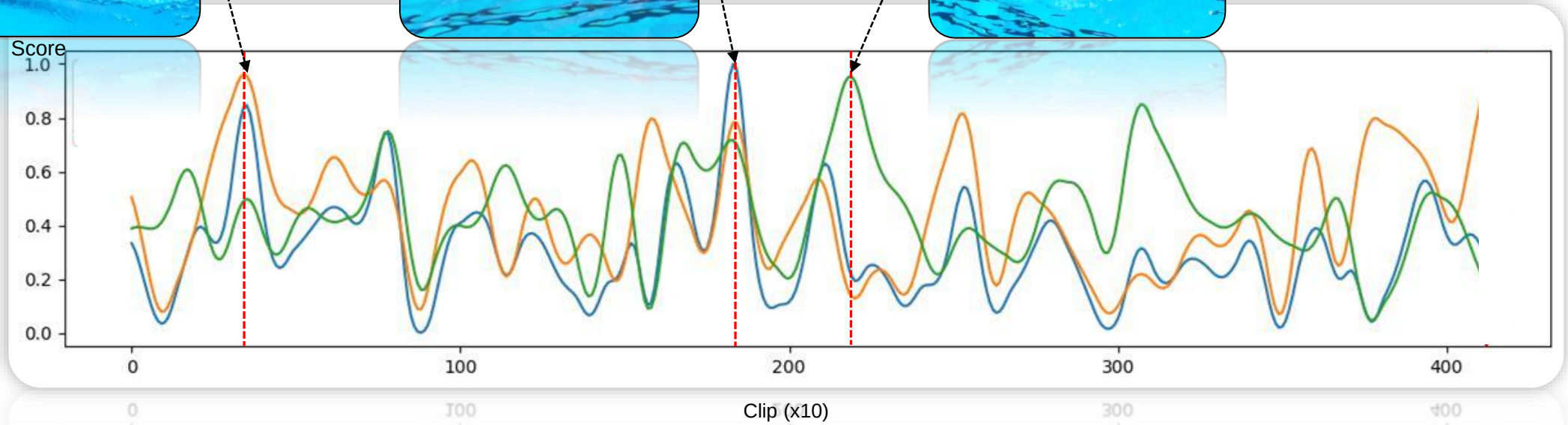
Weight: 0.50, Score: **0.96**
Overall: 0.84



Weight: 0.72, Score: 0.78
Overall: **0.98**

Weight: **0.95**, Score: 0.18
Overall: 0.24





Conclusion

- Explored interpretability in long-term AQA task.
- Proposed Attention Loss and Query initialization module to mitigate Temporal skipping problem.
- Proposed weight-score regression head for improve interpretability.
- Future work
  - Exploring **evaluation methods** the interpretability of AQA networks.
  - Other factor for evaluating action quality such as **artistic** quality.

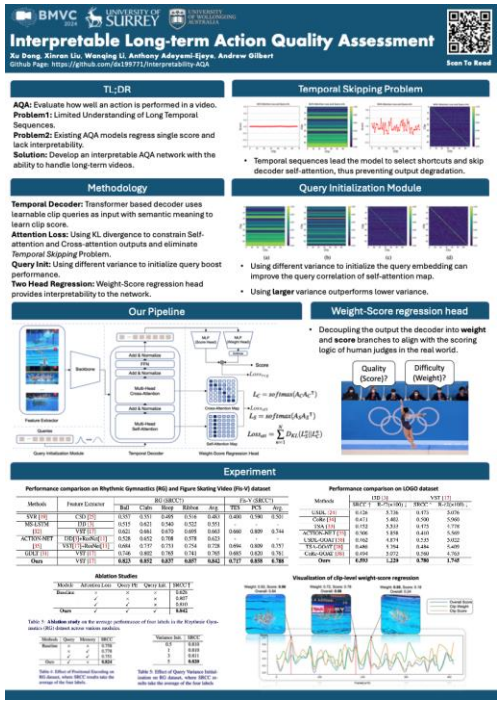
Thank you!

Interpretable Long-term Action Quality Assessment

Xu Dong[1], Xinran Liu[1], Wanqing Li[2], Anthony Adeyemi-Ejeye[1], Andrew Gilbert[1]

[1]University of Surrey, Guildford, UK
[2]Advanced Multimedia Research Lab, University of Wollongong, Wollongong, Australia

Find us at poster #517!



Read our paper

Find me on LinkedIn and WeChat!