FineCausal: A Causal-Based Framework for Interpretable Fine-Grained Action Quality Assessment

Ruisheng Han[1], Kanglei Zhou[2], Amir Atapour-Abarghouei[1], Xiaohui Liang[2], Hubert P. H. Shum[1]†
[1]Durham University [2]Beihang University

Background

- What is Action Quality Assessment (AQA)?
  – Automated scoring of human performance (e.g., dives, gymnastics) from video.
- Why it's challenging?
  – **Subtle motion cues** (body alignment, splash size) are buried in high-dimensional inputs.
  – **Context confounders** (lighting, camera angle, background clutter) bias predictions.
  – **Temporal dynamics**: each sub-action (take-off → somersault → entry) causally affects the next.

Motivation

- Robustness to spurious correlations by avoiding simplistic foreground–background fusion.
- Leverage both athlete and environment information in a principled way, rather than masking out or naively blending features.
- Move beyond a single black-box score to reveal which spatial regions and temporal stages most influence the final quality assessment.

Code

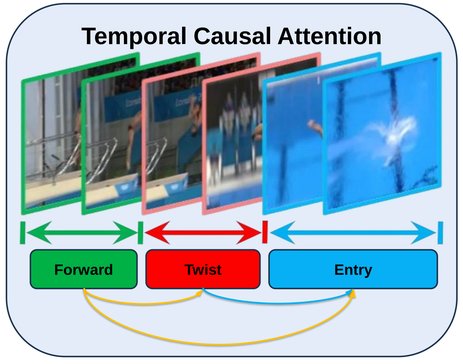

Figure 3. Temporal causal attention models the influence of each stage on the next.

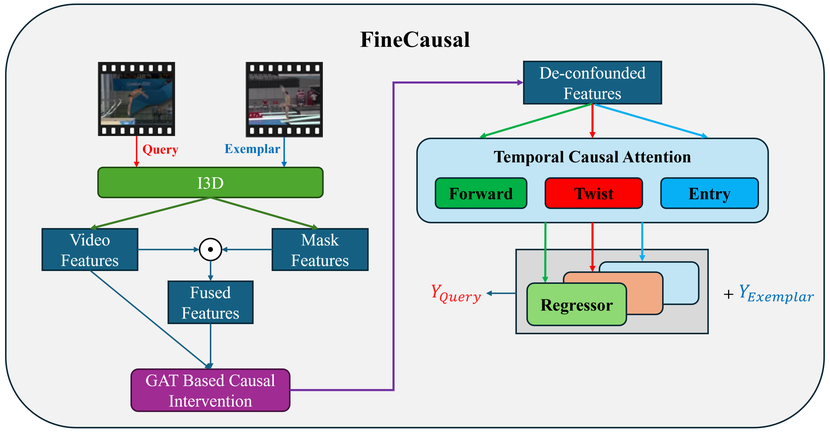

Figure 1. The architecture of FineCausal.

Contributions

- We introduce **FineCausal**, the first causal-based AQA framework that enhances interpretability by explicitly modelling the causal dependencies among video features.
- We propose a GAT-based intervention module that adaptively integrates both human-centric foreground cues and valuable background context.
- We develop a temporal causal attention module to capture fine-grained, stage-wise relationships across time.

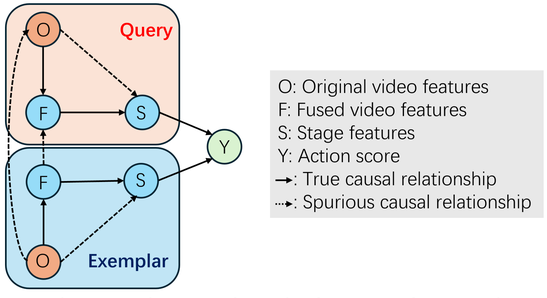

Figure 2. The causal graph of our AQA framework.

Experiments

Methods	AQA Metrics	
	$\rho\uparrow$	$R\text{-}\ell_2\downarrow(\times 100)$
C3D-LSTM [19]	0.6969	1.0767
C3D-AVG [18]	0.6801	0.6251
MSCADC [18]	0.7688	0.9372
I3D+MLP [22]	0.8227	0.4878
USDL [22]	0.8351	0.5104
MUSDL [22]	0.8240	0.4212
CoRe [31]	0.9308	0.3068
TSA [29]	0.9361	0.2746
HGCN [37]	0.9381	0.2321
FineParser [30]	0.9435	0.2602
FineCausal (Ours)	**0.9447**	**0.2338**

Methods	TAP Metrics	
	AIoU@0.5 ↑	AIoU@0.75 ↑
TSA [29]	0.9239	0.5007
FineParser [30]	**0.9946**	**0.9467**
FineCausal (Ours)	<u>0.9937</u>	<u>0.9453</u>

Table 1. Comparisons of performance with state-of-the-art AQA methods on the FineDiving-HM Dataset. The best result is highlighted in **bold**, and the second-best result is <u>underlined</u>.

Conclusion

- **FineCausal** disentangles and balances foreground and background cues while modelling fine-grained temporal dependencies.
- **Limitations:** Requires expert-specified causal structures and high-quality, fine-grained annotations, which may not be available for all sports or domains.
- Future directions:
  – Leverage semi-/self-supervision and automated annotation for broader applicability.
  – Extend the FineCausal framework to other fine-grained action assessment tasks.