

Causality-aware Multi-view Fusion and Human-Object Interaction Modeling for Action Quality Assessment in Freestyle Skiing

陳 詩玥 池永研究室 修士課程修了

Background

Judges-Only Scoring (Current Situation)

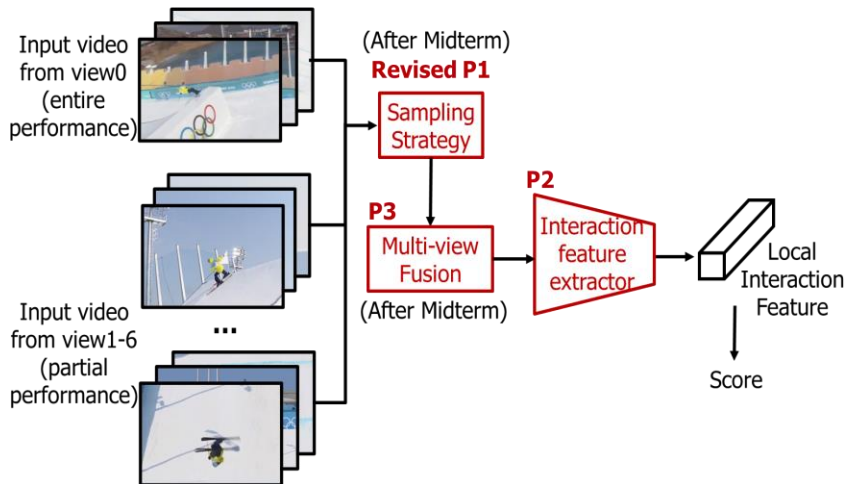


AQA Support Scoring (Highly Expected)



Target Design an Action Quality Assessment (AQA) system in Freestyle Skiing

Proposals



Global Feature(CW1)



Information from entire image

Local Feature(P2)



Information from human-object centric region

Reduce background influence irrelevant to scoring

Positional Feature(P2)



Relative position between human and object (interactive motion)

Both Local appearance and positional feature are adopted

Challenges

Challenge 1: Temporal Feature Selection (When to focus)



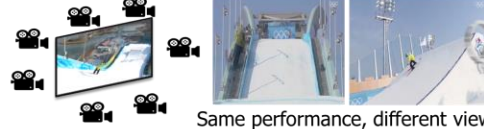
Hard to localize scoring-related frames

Challenge 2: Spatial Feature Selection (Where to look)



Hard to extract human-object features from entire scene

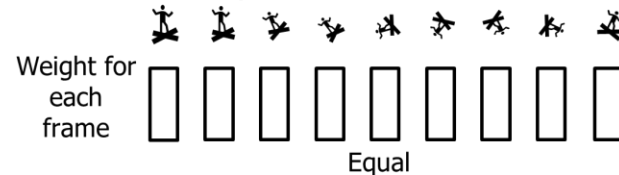
Challenge 3: View-aware Feature Selection (Which camera view)



Hard to quantify each view's contribution when combining them

Same performance, different views

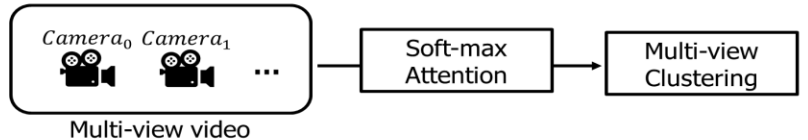
Uniform sampling



Entropy-based key frame sampling (Adopted)

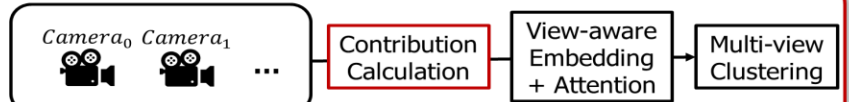


Multi-view Clustering without Causal Inference



Limitation: The method is black-box and fail to explain which view affects scoring.

Multi-view Clustering with Causal Inference



Merit: Based on each view's contribution, adjust attention module to give more weight to more scoring-related view.

Experiment Results

Method	RC ↑	RL2 ↓
CW1 ^[1]	0.6495	5.37
CW2 ^[2]	0.6699	5.59
USDL ^[3]	0.4944	6.36
CW1 ^[1] + P1	0.6733	5.13
CW1 ^[1] + P1+P2	0.6928	5.07
CW1 ^[1] + P1 + P2 + P3	0.7846	3.97

Conclusion

- The proposed method achieves a Spearman's Rank Correlation of 0.7846, and Relative L2-distance of 3.97, compared to baseline results of 0.6495 and 5.37 respectively

