Towards Debiasing Frame Length Bias in Text-Video Retrieval via Causal Intervention

Burak Satar [1,2] Hongyuan Zhu [1] Hanwang Zhang [2] Joo Hwee Lim [1,2]

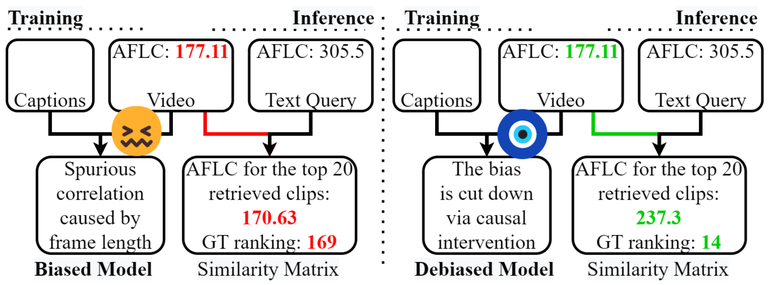
[1]Institute of Infocomm Research, A*STAR [2]SCSE, Nanyang Technological University

BMVC 2023

Introduction

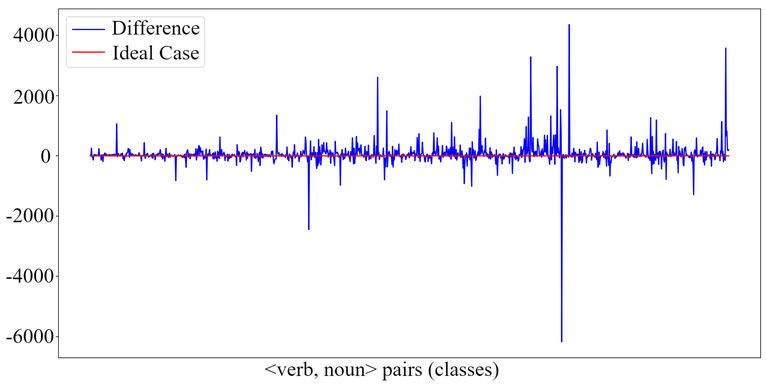
We present a systematic study on **a temporal bias** due to frame length discrepancy **between training and test sets** of trimmed video clips of EK-100, YC2 and MSRV-TT.

Problem.



Irrelevant clips are retrieved due to bias coming from the discrepancy. AFLC: Avg Frame Length of a Class.

Bias verification.

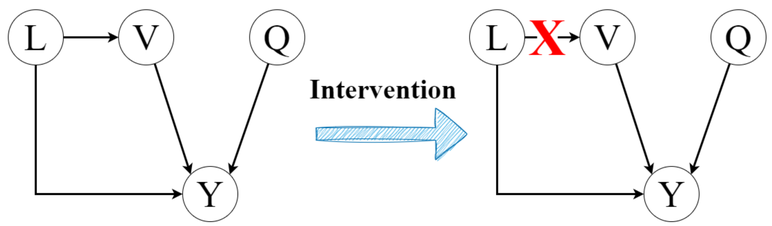


We calculate the **discrepancy among <verb, noun> pairs** (classes) by the average frame length difference between the training and test sets.

Debiasing Method

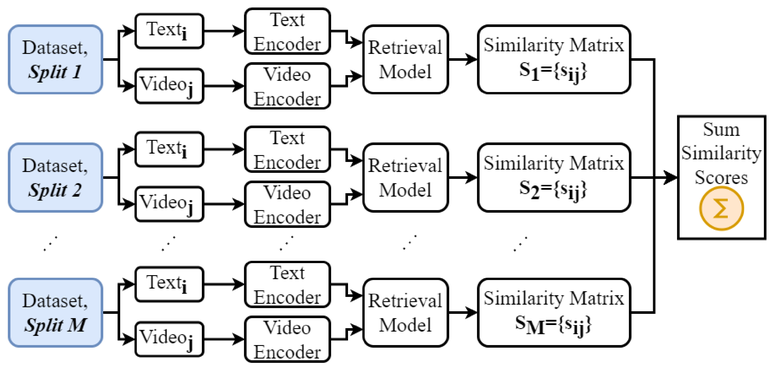
We propose a novel causal intervention method to remove this spurious correlation.

Structural causal model.



L –> *Y:* Natural effect on similarity matrix.
L –> *V:* Bias in videos, which **should be cut off.**

Model architecture.



Causal Intervention Formula.

$$E[Y|do(V,Q)] = \sum_{l} P(L=l|V,Q)E[Y|V,Q,L=l]$$
$$\triangleq \sum_{k=1}^{M} (L_k) f_k\{V,Q\}$$

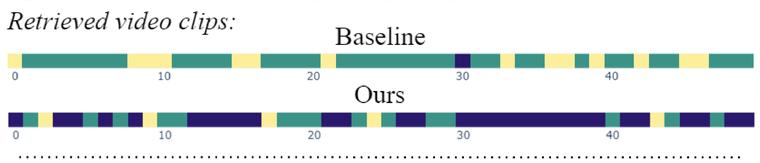
Results

Quantitative.

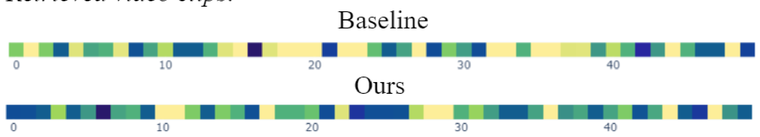
Method	nDCG			mAP		
	V2T	T2V	AVG	V2T	T2V	AVG
Epic-Kitchens-100						
Baseline	39.40	38.91	39.15	40.47	36.60	38.54
Baseline + RmvRand	39.69	38.42	39.06	40.37	35.7	38.04
Baseline + RmvAll	40.06	38.82	39.44	41.01	36.34	38.67
Baseline + Ensemble	40.38	39.15	39.76	43.17	38.80	40.98
Baseline + Ours	**42.73** (+3.33)	**40.61** (+1.70)	**41.67** (+2.52)	**45.36** (+4.89)	**37.80** (+1.20)	**41.58** (+3.04)

Qualitative.

Epic-Kitchen-100 | *Textual query:* pick up saucers



YouCook2 | *Textual query:* now add the beef 2 tbsp of flour 1 tsp of paprika 1 tbsp of tomato puree 2 bay leaves and 300ml beef stock



Contributions.

1) **The first** to verify the frame length bias,
2) **The first** to propose a **debiasing method** with causal inference **in text-video retrieval task**,
3) The nDCG metric shows that **the bias is mitigated**.

arXiv (pdf) Project Page