IEEE International Conference on Multimedia and Expo 2023
Brisbane Convention & Exhibition Centre
10-14 July 2023

Exploring Opinion-Unaware Video Quality Assessment with Semantic Affinity Criterion

ICME2023, Paper 766

Presented by Haoning Wu, Nanyang Technological University
11 Jul 2023, Brisbane

Background: In-the-wild VQA

What is the ultimate goal of robust VQA?

- In-the-wild VQA, a.k.a. real-world VQA, sometimes UGC-VQA.
- It is hard as it includes an open setting for VQA:
- - **open distribution***: a robust method should apply to any videos.*
- *-* **open audience:** *unbiased evaluation instead of reflecting particular preferences*
- *-* **open definition:** *should be able to evaluate respectively with specific requirements*

Background: In-the-wild VQA

How far are we towards this goal?

- At present, with limited scale of existing NR-VQA databases,
- The three open settings are less or more compromised:
- \- **open distribution**: *solely focusing on specific categories of contents*
- \- **open audience:** *opinions in different datasets are biased towards its protocols*
- \- **open definition:** *only a single "overall quality" score is provided*

Background: In-the-wild VQA

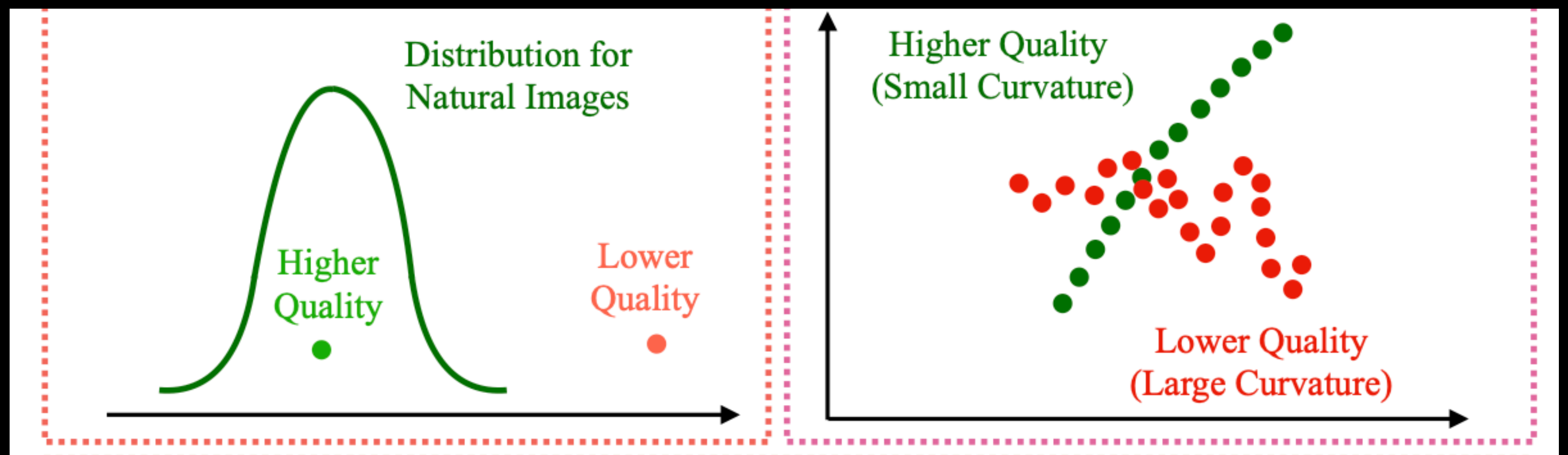
How far are we towards this goal?

- All these three issues hinder VQA to be generalizable (robust).
- For instance, state-of-the-art VQA methods **only** trained on KoNViD-1k are not well generalized into YouTube-UGC:
- VIDEVAL (0.443 PLCC), MDTVSFA (0.390 PLCC), DisCoVQA (0.447 PLCC)
- Not Enough Accurate on Close-Set Benchmarks.
- A long path towards real-world application on in-the-wild **open** settings.

Background: Zero-Shot VQA

Criterion-based Video Quality Assessment

NIQE TPQI



- We would like to explore how to evaluate video quality without VQA datasets.
- To conclude, they set a **CRITERION** between ‘high quality’ and ‘low quality’.
- For instance, **NIQE**, assumes that high quality visual contents follow specific distributions. Failing to fall into the distributions reflects low quality.
- For instance, **TPQI**, assumes that high quality videos should have frames with straight neural representations in the temporal dimension.

Background: Semantics in VQA

Semantic Criterion for Video Quality Assessment via CLIP

- However, they are not aware of semantic information, and may fall short on semantic-related quality comparisons, such as situations below:
- *(a) Is a blurry animal or a clear sky with better quality? (**Semantic Guidance**)*
- *(b) Is a nice flower or a dull scene with better quality? (**Semantic Preference**)*
- Both are hard to be answered without semantic awareness.
- Thus, we should build a **criterion** based on high-level information from videos.
- Via CLIP (Contrastive Language-Image Pretraining).

SAQI Index

Semantic Criterion for Video Quality Assessment via CLIP

- The proposed criterion, the **SAQI** index, can be considered as a soft classification between two description pairs:
- 1. Good *vs* Bad
- 2. High Quality *vs* Low Quality
- SAQI = average probability on the two positive descriptions

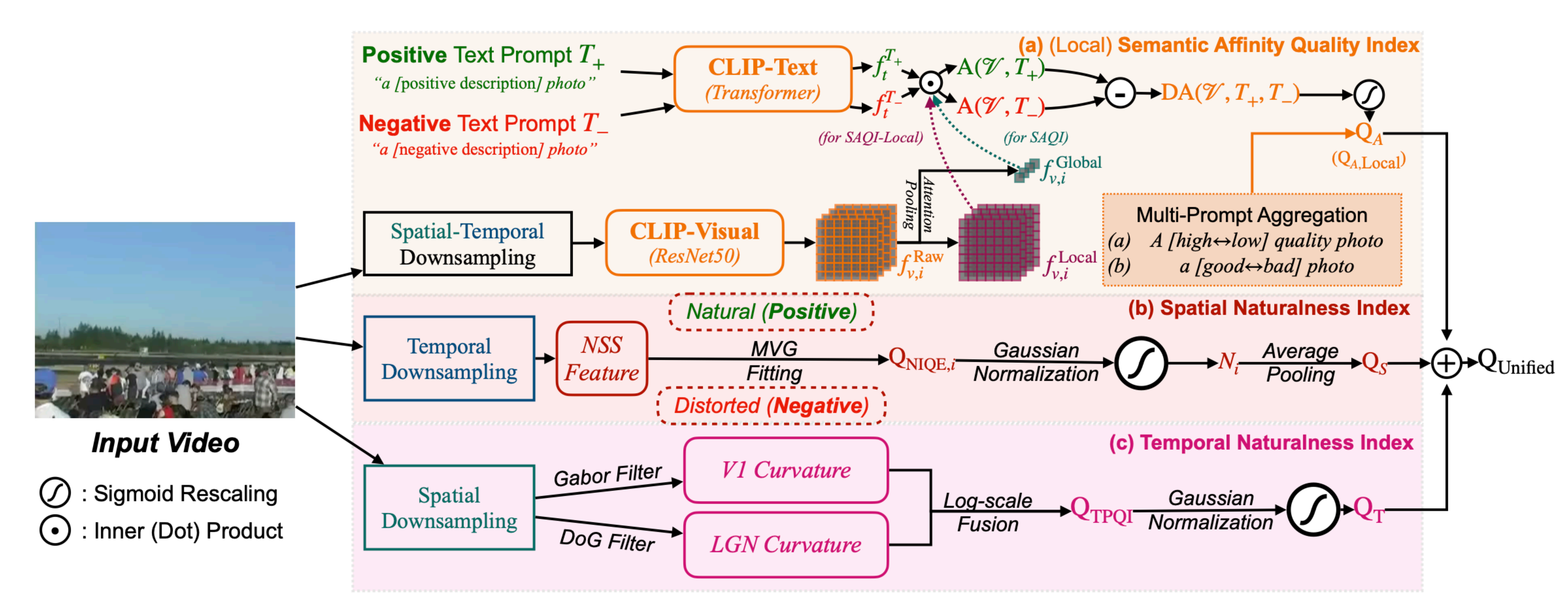
BVQI Index

Supplementing SAQI Index with traditional criterions

- Drawbacks of SAQI (due to the Internet-collected training scheme)
- *1. Lack of temporal modeling*
- *2. Weak modeling on traditional distortions (compression, transmission)*
- It can cooperate with traditional criteria to relieve the drawbacks.

BVQI Index

Method Overview



BVQI Index

Quantitative Results: *no-training > OOD-training*

Compare with Other Zero-shot

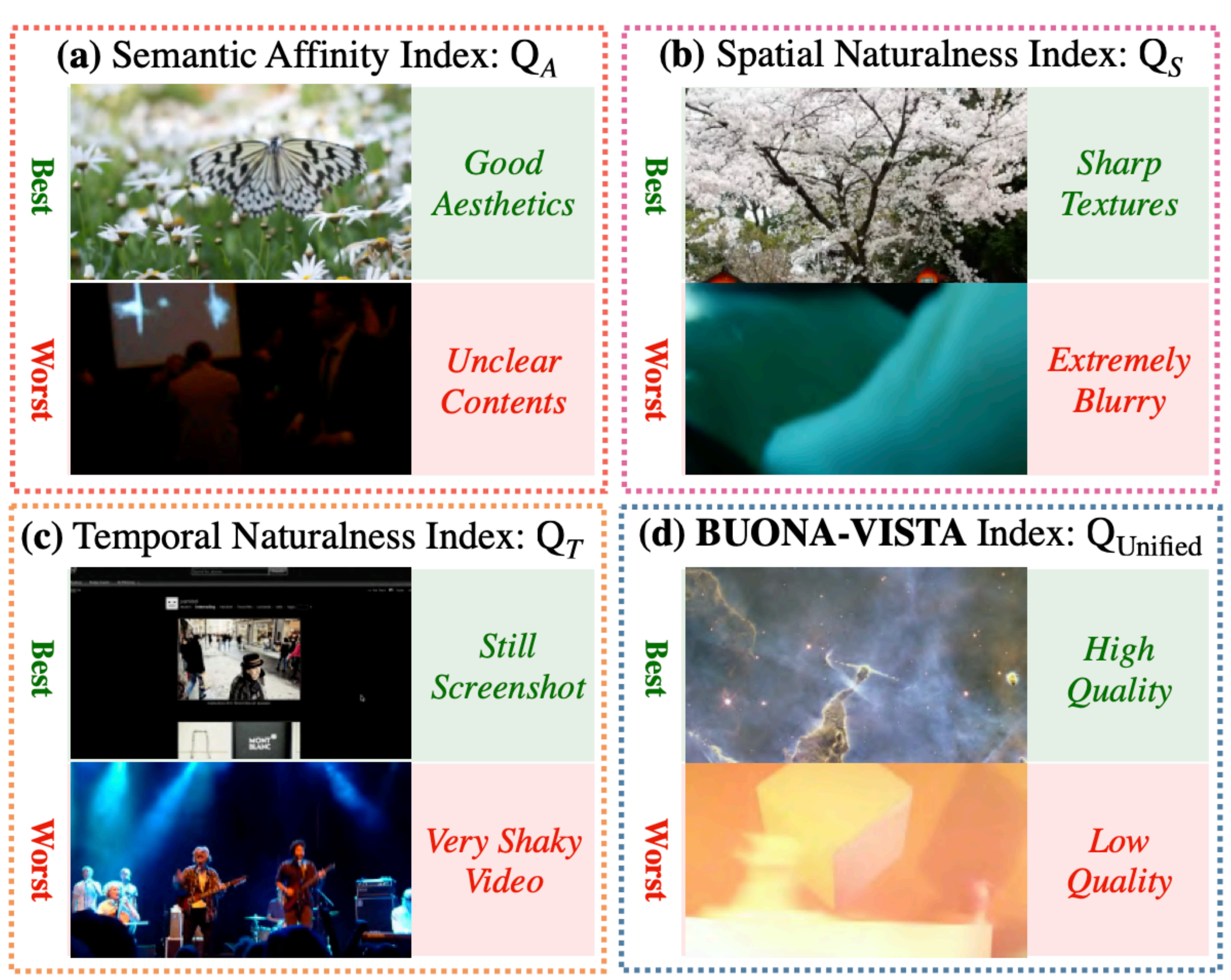
Dataset	LIVE-VQC		KoNViD-1k		YouTube-UGC		CVD2014	
Methods	SRCC↑	PLCC↑	SRCC↑	PLCC↑	SRCC↑	PLCC↑	SRCC↑	PLCC↑
(a) Zero-shot Quality Indices:								
(*Spatial*) NIQE (Signal Processing, 2013) [16]	0.596	0.628	0.541	0.553	0.278	0.290	0.492	0.612
(*Spatial*) IL-NIQE (TIP, 2015) [68]	0.504	0.544	0.526	0.540	0.292	0.330	0.468	0.571
(*Temporal*) VIIDEO (TIP, 2016) [34]	0.033	0.215	0.299	0.300	0.058	0.154	0.149	0.119
(*Temporal*) TPQI (ACMMM, 2022) [17]	0.636	0.645	**0.556**	0.549	0.111	0.218	0.408	0.469
(*Semantic*) SAQI (*Ours, ICME2023*)	0.629	0.638	0.608	0.602	**0.585**	**0.606**	0.685	0.692
(*Semantic*) SAQI-Local (*Ours, extended*)	**0.651**	**0.663**	**0.622**	**0.620**	**0.610**	**0.616**	**0.734**	**0.731**
(*Aggregated*) **BVQI** (*Ours, ICME2023*)	**0.784**	**0.794**	**0.760**	**0.760**	0.525	0.556	**0.740**	**0.763**
(*Aggregated*) **BVQI-Local** (*Ours, extended*)	**0.794**	**0.803**	**0.772**	**0.772**	**0.550**	**0.563**	**0.747**	**0.768**

Compare with Training-based **(OOD)**

Train on	KoNViD-1k				LIVE-VQC				Youtube-UGC			
Test on	LIVE-VQC		Youtube-UGC		KoNViD-1k		Youtube-UGC		LIVE-VQC		KoNViD-1k	
	SRCC↑	PLCC↑	SRCC↑	PLCC↑	SRCC↑	PLCC↑	SRCC↑	PLCC↑	SRCC↑	PLCC↑	SRCC↑	PLCC↑
TLVQM (2019, TIP) [3]	0.573	0.629	0.354	0.378	0.640	0.630	0.218	0.250	0.488	0.546	0.556	0.578
CNN-TLVQM (2020, MM) [7]	0.713	0.752	**0.424**	**0.469**	0.642	0.631	0.329	0.367	0.551	0.578	0.588	0.619
VIDEVAL (2021, TIP) [8]	0.627	0.654	0.370	0.390	0.625	0.621	0.302	0.318	0.542	0.553	0.610	0.620
MDTVSFA (2021, IJCV) [42]	**0.716**	**0.759**	0.408	0.443	0.706	**0.711**	**0.355**	**0.388**	**0.582**	**0.603**	**0.649**	**0.646**
GST-VQA (2022, TCSVT) [6]	0.700	0.733	NA	NA	**0.709**	0.707	NA	NA	NA	NA	NA	NA
BVQI-Local (*before fine-tuning*)	0.794	0.803	0.550	0.563	0.772	0.772	0.550	0.563	0.794	0.803	0.772	0.772

BVQI Index

Qualitative Results: *best/worst*



Conclusion

A Step Towards Open-World in-the-wild VQA

- We move a step forward to ‘open-setting’ in-the-wild VQA.
- This is achieved by building **CRITERION** on different aspects of visual quality:
- - Semantics (Guidance, Preference)
- - Spatial Traditional Distortions
- - Temporal Naturalness
- We hope this method can be a well-applicable method in real-world.,

Follow our updates!

- *Links for BVQI (or a longer name, BUONA-VISTA)*

ICME Paper on Arxiv

Extended Paper on Arxiv

Code Repository

Key Features:

- Robustly evaluate video quality **without training** from any MOS scores.
- **Localized** semantic quality prediction.
- Given a small set of MOS-labelled videos, can **robustly+efficiently fine-tune** on it.

- *Related Links for Our Team (VQAssessment Team at NTU-Singapore)*

GitHub Homepage

***Supervised** VQA State-of-the-Arts:*

*FAST-VQA (2022, **146** Stars)*

*DOVER (2023, **85** Stars)*