A Comparative Evaluation of Temporal Pooling Methods for Blind Video Quality Assessment

Zhengzhong Tu[1⋆], Chia-Ju Chen[1⋆], Li-Heng Chen[1], Neil Birkbeck[2], Balu Adsumilli[2], and Alan C. Bovik[1]

[1]The University of Texas at Austin, [2]YouTube Media Algorithms Team, Google Inc.

Presenter: Chia-Ju (Christie) Chen

Check out our paper @SMR-05.8

The University of Texas at Austin LIVE Laboratory for Image & Video Engineering

AbuDhabi ICIP 2020

YouTube

Agenda

- Background
- Temporal Pooling Methods
  - Traditional means
  - Means emphasizing low-quality parts
  - Means emphasizing memory effects
- Ensemble Temporal Pooling
- Experimental Results
- Observations and Recipe

Background

- **Blind Video Quality Assessment (BVQA):**
  Blind predicting perceptual quality of a video clip
  - Video-level features + Regression
  - Blind frame quality assessment (BIQA) + temporal quality pooling
- **BIQA + Temporal Pooling**
  - Simple but effective
  - Evidence from subjective experiments [1]
  - Easily extensible to build future BVQA based on future BIQA models

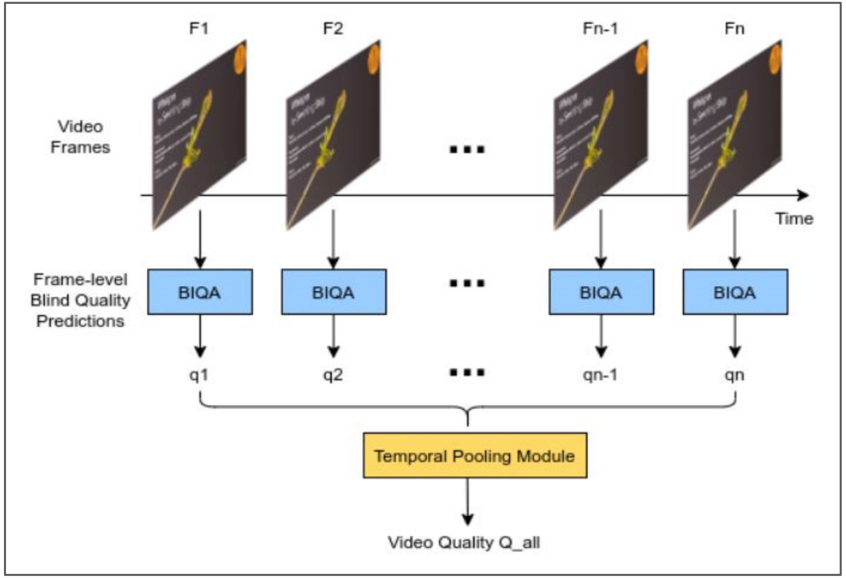

Fig 1. Building BVQA models by temporal pooling of BIQA-predicted scores

[1] S., Kalpana, and A. C. Bovik. "Temporal hysteresis model of time varying subjective video quality." *ICASSP*, 2011.

Temporal Pooling Methods

- **Traditional Means**
  - Arithmetic Mean:
    $$Q = \frac{1}{N}\sum_{n=1}^{N} q_n$$
  - Harmonic Mean:
    $$Q = \left(\frac{1}{N}\sum_{n=1}^{N} q_n^{-1}\right)^{-1}$$
  - Geometric Mean:
    $$Q = \left(\prod_{n=1}^{N} q_n\right)^{1/N}$$
  - Lp Minkowski Mean:
    $$Q = \left(\frac{1}{N}\sum_{n=1}^{N} q_n^{p}\right)^{1/p}$$

N: number of frames
q_n: predicted quality of n-th frame
Q: final predicted video quality

Temporal Pooling Methods

- **Means emphasizing low-quality parts**
  - Percentile: $$Q = \frac{1}{|P_{\downarrow k\%}|} \sum_{n \in P_{\downarrow k\%}} q_n$$ where $P_{\downarrow k\%}$ denotes the set of lowest $k\%$ scores
  - VQPooling [2]: $$Q = \frac{\sum_{n \in G_L} q_n + w \cdot \sum_{n \in G_H} q_n}{|G_L| + w \cdot |G_H|}$$ Where G_L and G_H are low quality and high quality groups separated by k-means

    where $$w = \left(1 - \frac{M_L}{M_H}\right)^2$$
- Temporal Variation $$Q = \frac{1}{|P_{\uparrow k\%}|} \sum_{n \in P_{\uparrow k\%}} |q_n - q_{n-1}|$$

  where $P_{\uparrow k\%}$ is the set of largest $k\%$ absolute quality differences

[2] Park, J., Seshadrinathan, K., Lee, S., & Bovik, A. C. (2012). Video quality pooling adaptive to perceptual distortion severity. *TIP*.

Temporal Pooling Methods

- **Memory effects:**
  - Primacy and recency effects:

$$Q = \sum_{n=1}^{N} w_n q_n$$

Primacy:

$$w_n = \frac{\exp(-\alpha_p n)}{\sum_{k=0}^{L} \exp(-\alpha_p k)}, \ 0 \leq n \leq L$$

Recency:

$$w_n = \frac{\exp(-\alpha_r (L-n))}{\sum_{k=0}^{L} \exp(-\alpha_r (L-k))}, \ 0 \leq n \leq L$$

- **Hysteresis pooling** [3]

$$l_n = \begin{cases} q_n, & n = 1 \\ \min\limits_{k \in \mathcal{K}_{prev}} \{q_k\}, & n > 1 \end{cases}$$

$$\boldsymbol{v} = sort(\{q_k\}), \ k \in \mathcal{K}_{next}$$

$$m_n = \sum_{j=1}^{J} v_j w_j, \ J = |\mathcal{K}_{next}|$$

$$q'_n = \alpha m_n + (1-\alpha) l_n$$

$$Q = \frac{1}{N} \sum_{n=1}^{N} q'_n$$

[3] Seshadrinathan, K., & Bovik, A. C. Temporal hysteresis model of time varying subjective video quality. *ICASSP, 2011.*

Ensemble Temporal Pooling

- **We propose an ensemble-based pooling:**

$$Q_{\text{EPooling}} = \mathcal{F}(\boldsymbol{Q}),\ \boldsymbol{Q} = \{Q_i\},\ i = 1, 2, ..., I$$

where, Q is singly pooled score, and F() is a learned regression mapping.

- **We empirically selected three pooling methods to ensemble:**

  - Mean, VQPooling, Hysteresis

- **Why ensemble?**
  - More robust
  - Better performance on average

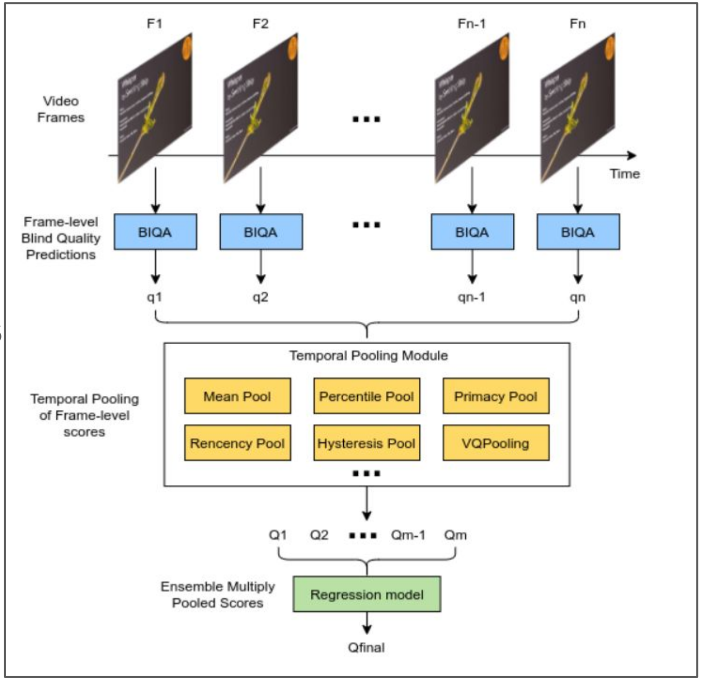

Fig 2. Ensenble multiply temporal pooling methods

Experiments

- **Selected 5 popular BIQA models:**
  - NIQE, BRISQUE, GM-LOG, HIGRADE, CORNIA
- **UGC video quality benchmarks:**
  - KoNViD-1k, LIVE-VQC
- **Parameters:**
  - p = 2 for Minkowski
  - k = 10% for percentile
  - $(L, \alpha_p, \alpha_r) = (180, 0.01, 0.01)$ for memory effects
  - $(\tau, \alpha) = (60, 0.8)$ for temporal hysteresis
- **Evaluation protocols:**
  - 100 iterations of 80%-20% train-test splits, report median SRCC/PLCC

Experimental Results

Table 1: Performance comparison of temporal pooling methods as evaluated on KoNViD-1k [34] and LIVE-VQC [35]. Each cell shows the median evaluation results formatted as SRCC/PLCC. The three best results along each column are **boldfaced**.

Database	KoNViD-1k					LIVE-VQC				
Pool/Model	NIQE	BRISQUE	GMLOG	HIGRADE	CORNIA	NIQE	BRISQUE	GMLOG	HIGRADE	CORNIA
Mean	0.552/**0.560**	0.673/0.676	0.662/0.671	0.690/0.696	**0.749/0.764**	0.600/0.631	0.597/0.632	0.575/0.618	0.532/0.570	0.694/0.743
Median	0.543/0.554	0.667/0.670	0.657/0.666	0.680/0.689	**0.750/0.760**	0.584/0.618	0.577/0.619	0.558/0.602	0.521/0.559	0.687/0.744
Harmonic	0.550/**0.560**	**0.674**/0.676	0.667/0.674	0.693/0.699	0.696/0.696	0.607/0.637	0.605/0.636	0.585/0.620	0.537/0.575	**0.709**/0.737
Geometric	0.551/**0.560**	**0.676/0.679**	0.666/0.673	0.692/0.698	0.747/**0.760**	0.604/0.634	0.600/0.631	0.578/0.617	0.537/0.573	0.698/**0.746**
Minkowski	0.552/0.559	0.672/0.676	0.661/0.670	0.689/0.695	0.736/0.746	0.597/0.628	0.596/0.630	0.574/0.615	0.538/0.569	0.688/0.739
Percentile	0.545/0.547	0.655/0.647	**0.674/0.678**	0.685/0.687	0.696/0.700	**0.630**/0.634	**0.629/0.647**	**0.606/0.627**	**0.586/0.610**	**0.712**/0.744
VQPooling	0.549/0.554	0.670/0.665	**0.672**/0.674	**0.698/0.701**	0.743/0.758	**0.628/0.644**	**0.617/0.658**	**0.605/0.633**	0.563/0.597	0.700/**0.753**
Variation	0.347/0.328	0.348/0.338	0.509/0.511	0.434/0.444	0.240/0.303	0.507/0.476	0.470/0.463	0.495/0.488	0.474/0.482	0.567/0.609
Primacy	0.541/0.552	0.668/0.671	0.647/0.653	0.684/0.690	0.726/0.741	0.601/0.631	0.573/0.627	0.575/0.613	0.535/0.561	0.684/0.737
Recency	**0.553**/0.558	0.670/0.667	0.660/0.667	0.690/0.694	0.745/0.754	0.584/0.615	0.586/0.626	0.561/0.599	0.518/0.555	0.670/0.729
Hysteresis	**0.563/0.569**	**0.684/0.681**	**0.681/0.684**	**0.703/0.707**	0.732/0.735	0.621/**0.638**	**0.621/0.650**	**0.600/0.629**	**0.570/0.595**	**0.711/0.756**
EPooling	**0.572/0.579**	0.670/**0.679**	0.670/**0.676**	**0.698/0.704**	**0.749/0.762**	**0.623/0.645**	**0.617**/0.646	**0.605**/0.623	**0.582/0.601**	0.705/0.743

Observations and Recipe

- Efficacy of temporal pooling is content-dependent:
  - If videos contain **more motion** or **temporal quality variation**, we recommend low-quality emphasizing pooling strategies like **Perceptile, VQPooling, or Hysteresis**.
  - If videos have less (camera) motion, like UGC videos uploaded to Flickr or YouTube, traditional sample mean may be adequate.
- Our proposed Ensemble pooling is an effective way to compensate between different pooling methods, thus delivering a more robust result.

Thanks for listening!

The University of Texas at Austin
LIVE
Laboratory for Image & Video Engineering

AbuDhabi ICIP 2020

YouTube