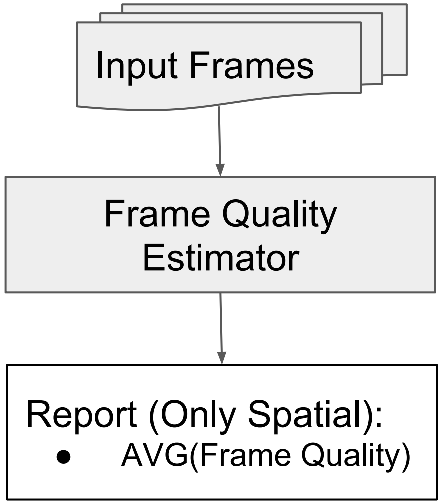
Quality Analysis For UGC Videos

Yilin Wang, YouTube Media Algorithms team

Video Quality Analysis

- Millions of User Generated Contents (UGC) are uploaded to YouTube everyday
  - Video Compression is critical
- Quality analysis is important for compression/transcoding
  - Popular quality metrics: PSNR, SSIM, VMAF, …
- Traditional video analysis framework
  - evaluate (reference) spatial quality issues for each frame
  - aggregate summary statistics (e.g. mean or worst 5%) of quality score per frame to an overall measure



Shortcomings of Traditional Framework

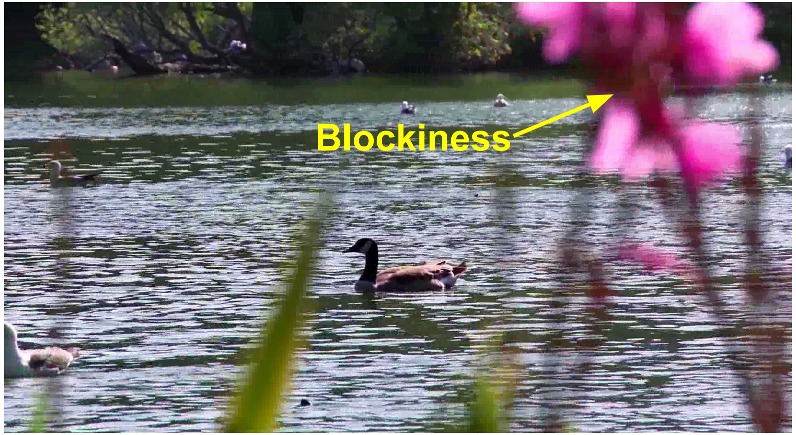
- Non-pristine uploaded version

Initial Raw Video

Uploaded Version

Shortcomings of Traditional Framework

- Hard to evaluate “positive” quality changes

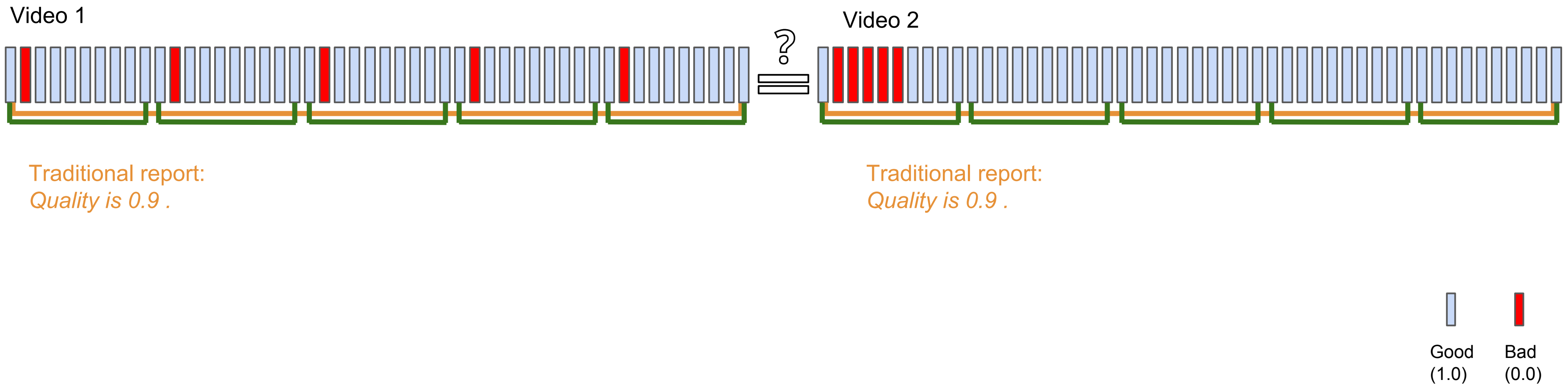


Uploaded

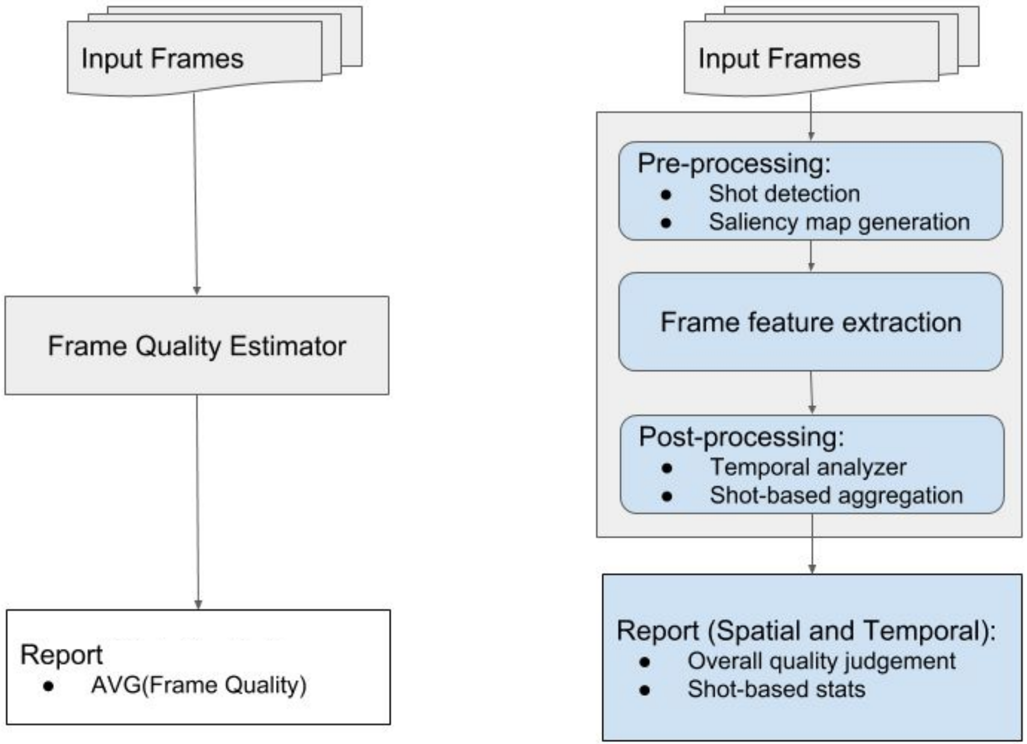
Transcoded

Shortcomings of Traditional Framework

- Ambiguous frame quality aggregation.



Video Quality Analysis Framework



Traditional Quality Analysis Framework

New Quality Analysis Framework

Preprossing

- Shot detection
  - judged by the difference between color histograms of previous and current frames
  - more reliable to assume consistent quality within the same shot instead of the entire video

- Saliency map generation
  - reweight impact of frame pixels

Frame Feature Extraction

- Goal
  - to extract useful features for further quality analysis in post-processing step.
  - mainly focusing on “describable” artifacts (e.g. banding, noise, blockiness, blur, ...)
- Novel non reference features
  - spatial: banding, noise, sleeq, ...
  - temporal: jerkiness, ...

Spatial Feature: Banding

Yilin Wang, Sang-Uok Kum, Chao Chen, Anil Kokaram, “A perceptual visibility metric for banding artifacts,” IEEE International Conference on Image Processing, 2016.

Banding Artifacts

Original

Transcoded

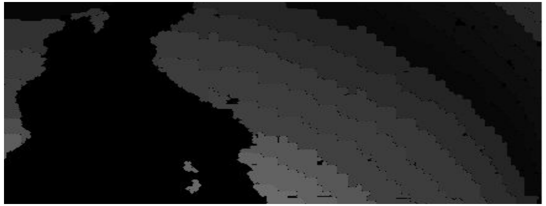
MOS (Mean Opinion Score): 40.5

	Metric Score	Predicted MOS
PSNR	*48.9811*	*97.7*
SSIM	*0.9904*	*100.0*
VMAF	*95.52*	*95.52*
Our Banding	*10.2233*	***51.1***

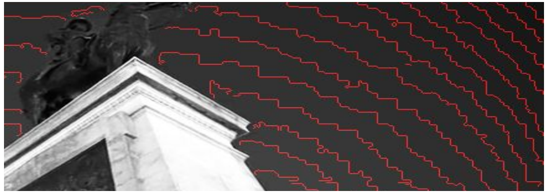
Banding Feature

- Uniseg
  - a large segment of pixels with same intensity
- Banding Edge
  - boundary pixels between two unisegs with close intensity

Input Frame



Unisegs



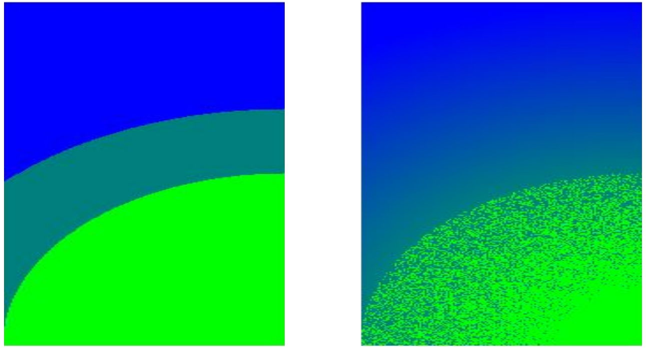
Detected Banding Edges
(clean and tightly matching with the visible bandings)

Edges (Detected by Canny)

Edge Coherence

- Intensity contrast outside the banding edge

$$\text{Edge Coherence} = 1 - \min\left(1, \frac{\text{outside pixels with the same intensity as the banding edge}}{\text{outside pixels with different intensities as the banding edge}} \right)$$



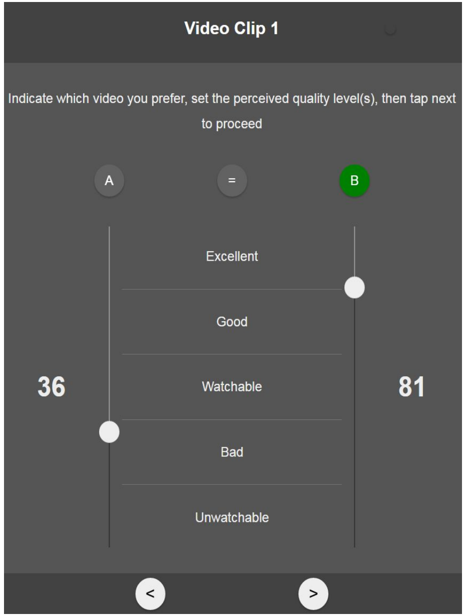
Edge Coherence: High v.s. Low

Banding Score

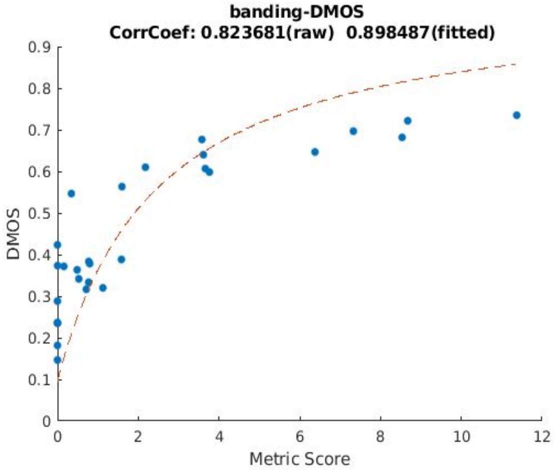
$$\text{banding} = \sum_i \left((\text{edge}(i).\text{length} \;/\; \text{diagonal\_length}) * (\text{edge}(i).\text{coherence} > \text{T})\right)$$

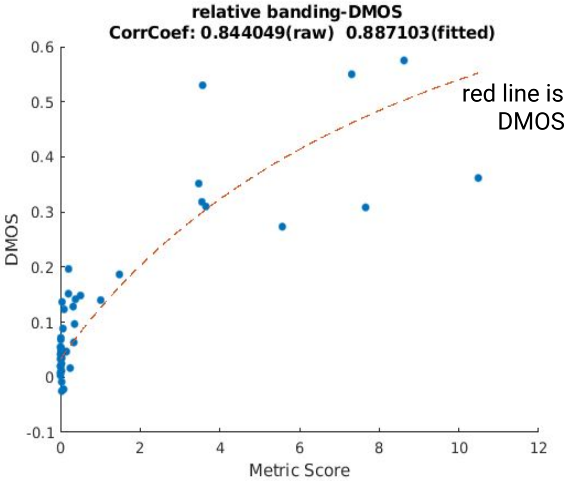
Subjective Experiment

- 8 original videos, each video is used to generate 3 test samples
- 1 hour test
- 25 participants



Correlation with Subjective Banding Scores





red line is fitted by a logistic model:

DMOS = 1 − 1/(c0 + exp(c1 · metric_score + c2))

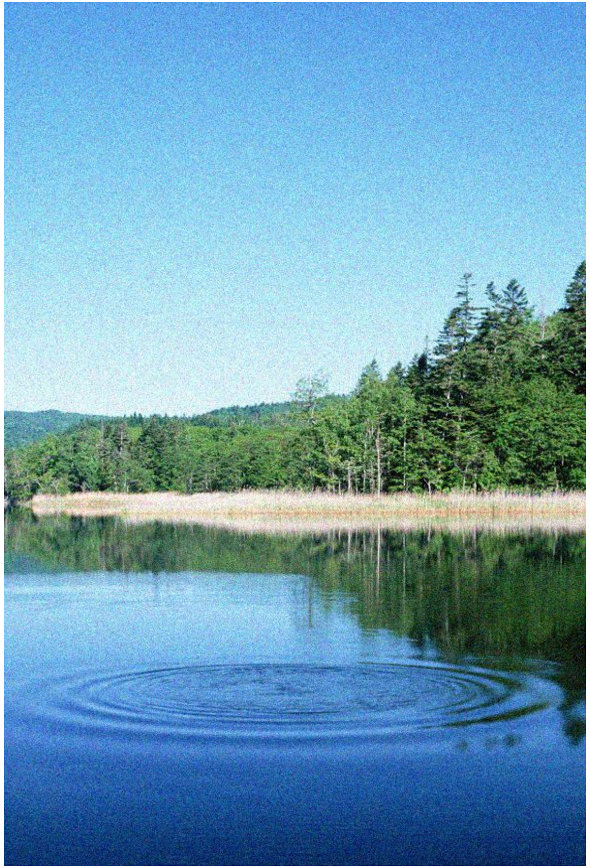
	PSNR	SSIM	VMAF	**Banding**
No ref	n/a	n/a	n/a	0.892
Ref	0.512	0.353	0.141	0.883

Spatial Feature: Noise

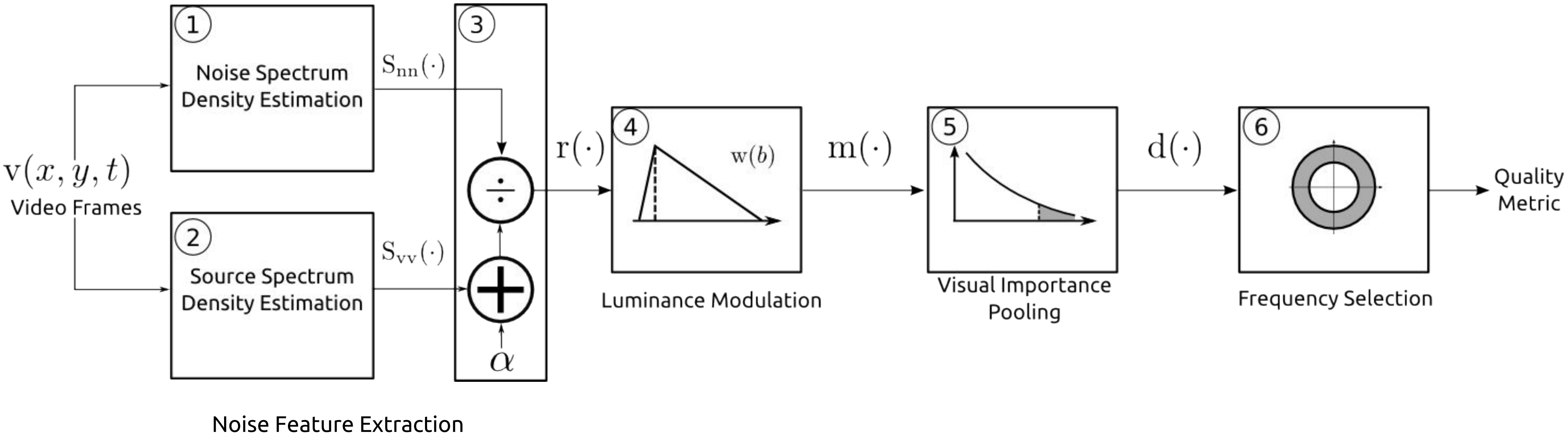
Chao Chen, Mohammad Izadi, and Anil Kokaram, "A no-reference perceptual quality metric for videos distorted by spatially correlated noise," ACM Multimedia, 2016.

Noise Matters

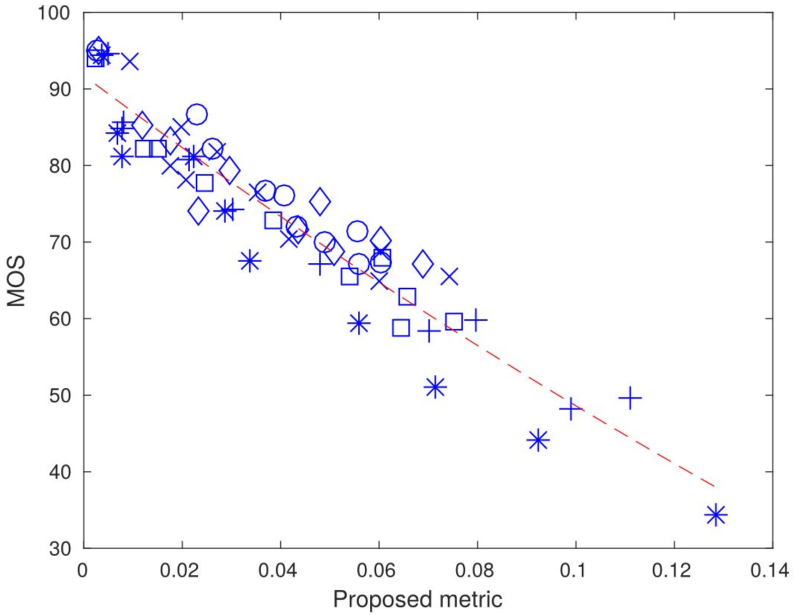
- Noise is inherent in uploaded videos
- Propagate through the video processing pipeline
  - Cause encoding artifacts
  - Waste encoding bits
- Need a metric to detect and measure noise
  - Detect noisy videos and Apply denoiser
  - Evaluate quality of denoised videos
  - Monitor quality of uploaded videos



Proposed Noise Metric



Performance Evaluation



Metrics	Linear Corr	Rank Corr	Prediction Error (VoR)
Noise	***0.9417***	***0.9545***	***16.8952***
PSNR	*0.7019*	*0.6549*	*86.6038*
SSIM	*0.8390*	*0.8103*	*65.0248*
VQM	*0.7450*	*0.7108*	*89.3259*
STMAD	*0.7100*	*0.7236*	*87.9041*

Correlation between subjective scores and proposed noise metric

Spatial Feature: Self-reference based LEarning-free Evaluator of Quality (SLEEQ)

Deepti Ghadiyaram, Chao Chen, Sasi Inguva, Anil Kokaram, "A no-reference video quality predictor for compression and scaling artifacts," IEEE International Conference on Image Processing, 2017

YouTube

Natural Scene Statistics

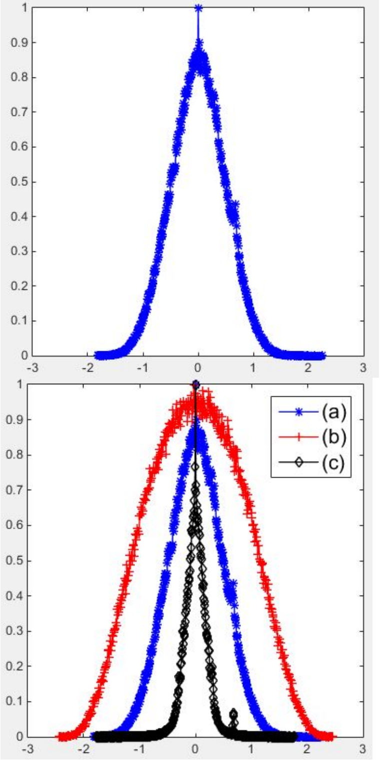
Divisive normalized pixel values of natural scenes follow Gaussian distribution.

Divisive Normalization

$$Y = \frac{I(i,j) - \mu(i,j)}{\sigma(i,j) + C}$$

Divisive Normalization

$$Y = \frac{I(i,j) - \mu(i,j)}{\sigma(i,j) + C}$$



Histogram of **Y**

D. L. Ruderman, “The statistics of natural images,” Netw., Comput. Neural Syst., vol. 5, no. 4, 1994

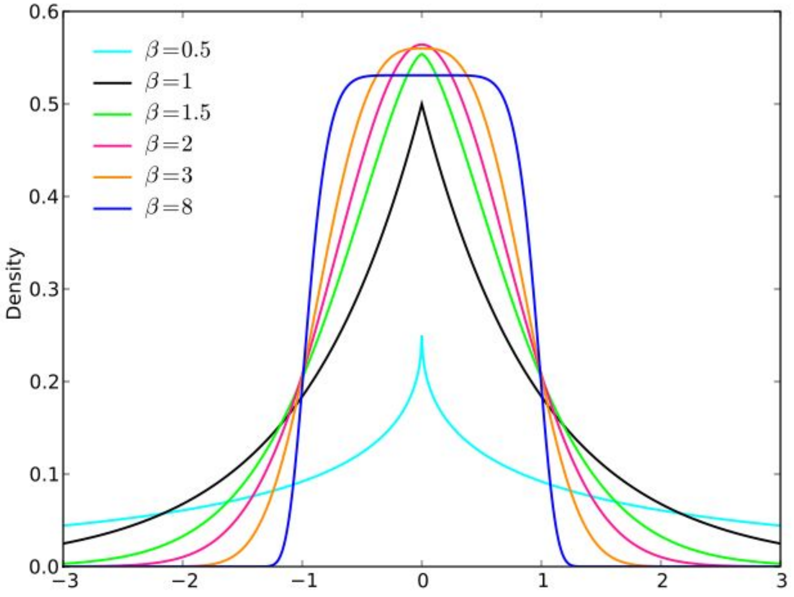
GGD Parameter

- Generalized Gaussian Distribution (GGD)

$$\frac{\beta}{2\alpha\Gamma(1/\beta)} \exp\left(-\left(\frac{|x-\mu|}{\alpha}\right)^{\beta}\right)$$

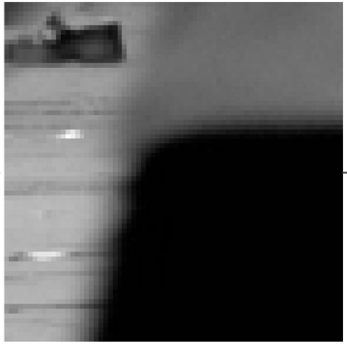
$$\beta = s\sqrt{\frac{\Gamma(1/\alpha)}{\Gamma(3/\alpha)}} \qquad \Gamma(a) = \int_0^\infty t^{a-1}e^{-t}dt \quad a > 0.$$

- ***β*** = 2 indicates good quality
- |***β*** - 2| is used as quality indicator

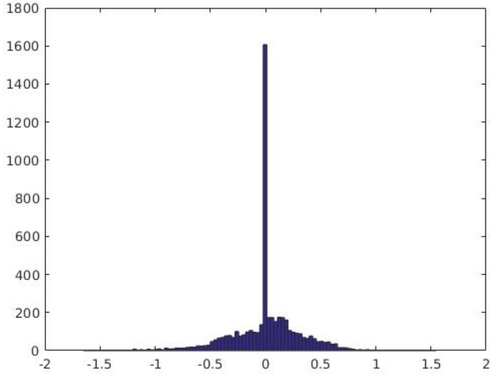


Spatial Complexity

- Natural Scenes Statistics does not apply to flat regions
- Calculate local variance $\boldsymbol{\sigma}$
- Skip Flat Blocks
  - Skip block with mean($\boldsymbol{\sigma}$) $\leqslant T_1$
  - Skip block with |mean($\boldsymbol{\sigma}$) - median($\boldsymbol{\sigma}$)| $\geqslant T_2$



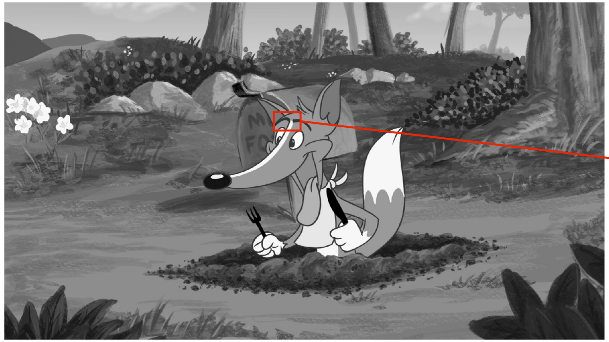
Block with flat region



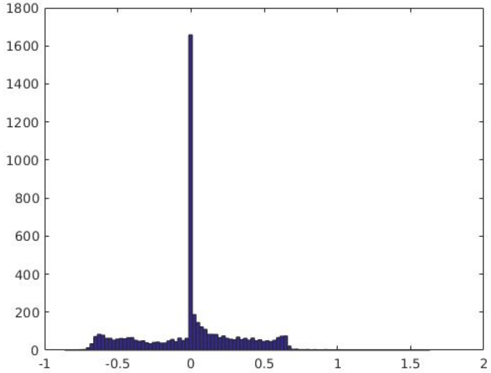
High spike around zero

Edge Strength

- Natural Scenes Statistics does not apply to strong edges
- Detect edges using Canny detector
- Skip blocks with strong edges



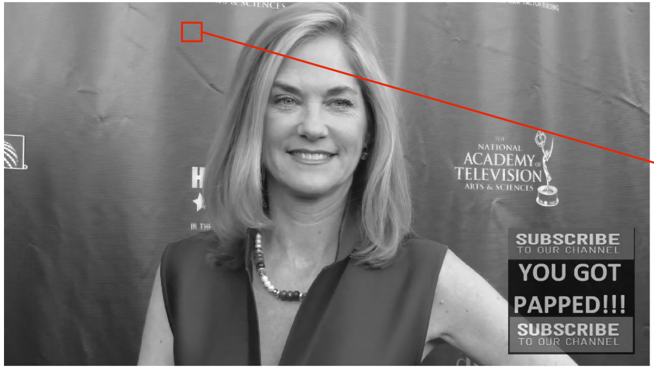
Block with strong edges

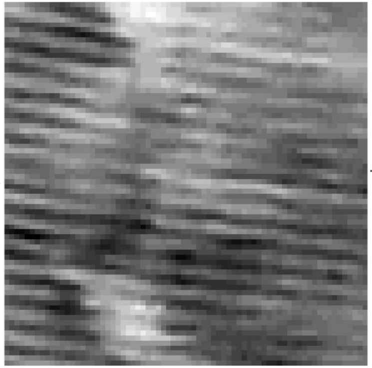


Asymmetric distribution

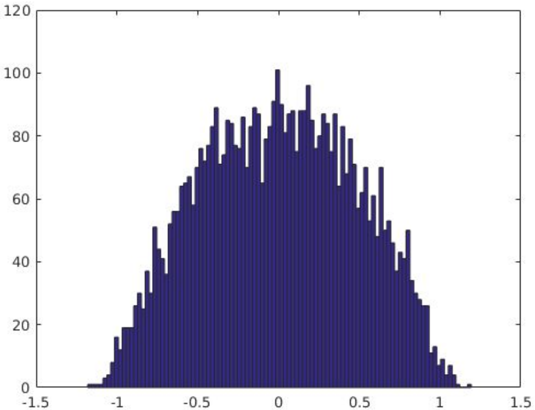
Texture Strength

- Natural Scenes Statistics does not apply to textures
- Detect texture using power spectrum density (PSD)
- Skip blocks with strong textures





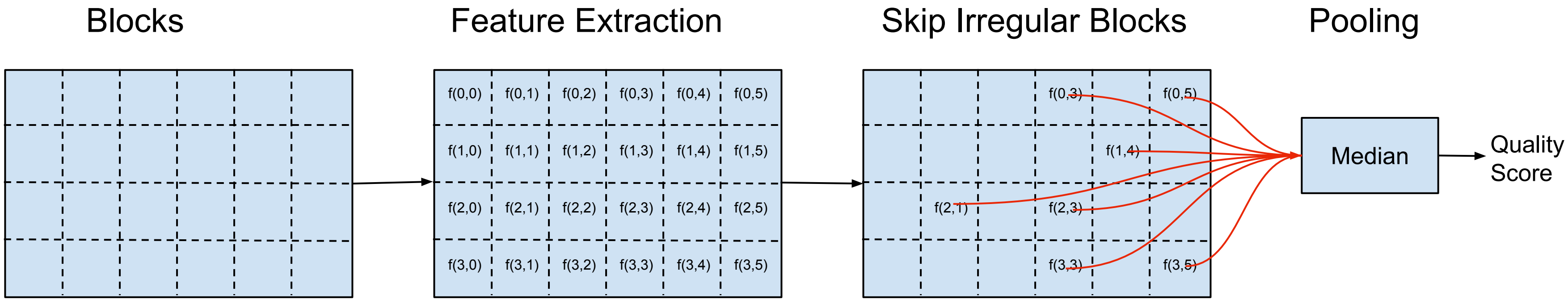
Block with flat region



Distribution Deviates From Gaussian

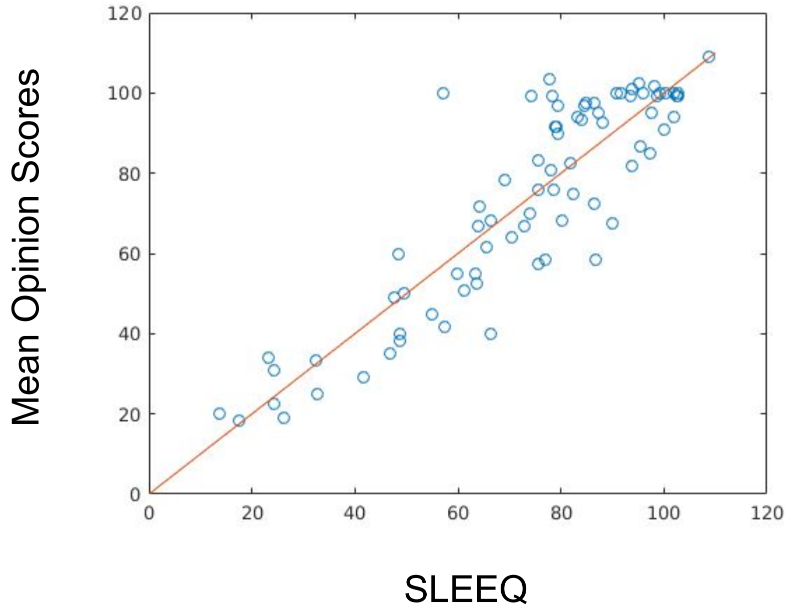
SLEEQ Algorithm

- Block-wise feature extraction + Irregular Block Removal



Performance

- On a known visual quality database
  - 79 videos
  - H264 compression artifacts
  - Upscaling artifacts
- SLEEQ performance
  - Linear Correlation 0.8915
  - MSE 11.7457 at scale [0, 100]



Temporal Feature: Jerkiness

Yilin Wang, Balu Adsumilli, "Video Quality Analysis Framework For Spatial and Temporal Artifacts," Applications of Digital Image Processing XLI, SPIE Optical Engineering + Applications, 2018

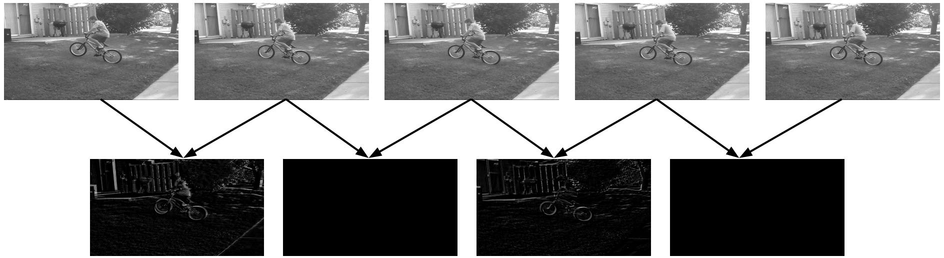
Jerkiness Artifacts

Jerkiness artifacts

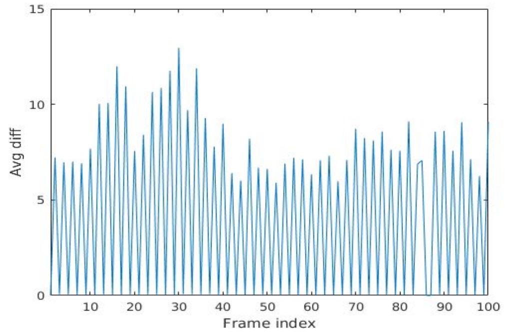
No jerkiness artifacts

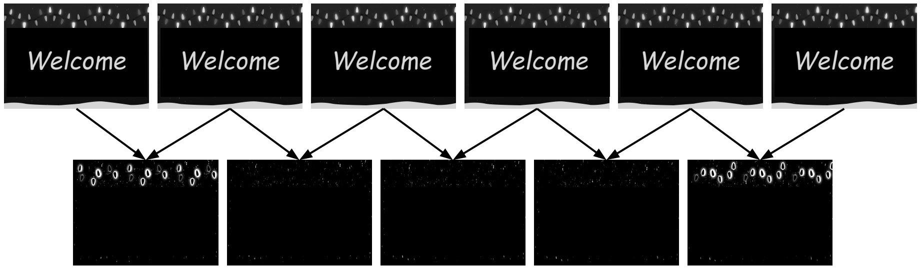
Jerkiness is a typical video artifacts caused by video compression/transcoding, especially when downsampling HFR (High Frame Rate) videos with insufficient sampling rates.

Frame differences for true and fake jerkiness artifacts

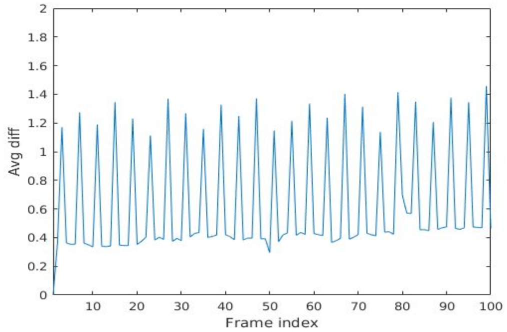


Avg diff: 13.41 0.06 9.80 0.06





Avg diff: 1.17 0.36 0.35 0.35 1.27



Absolute differences between neighboring frames.

Absolute differences for 100 consecutive frames.

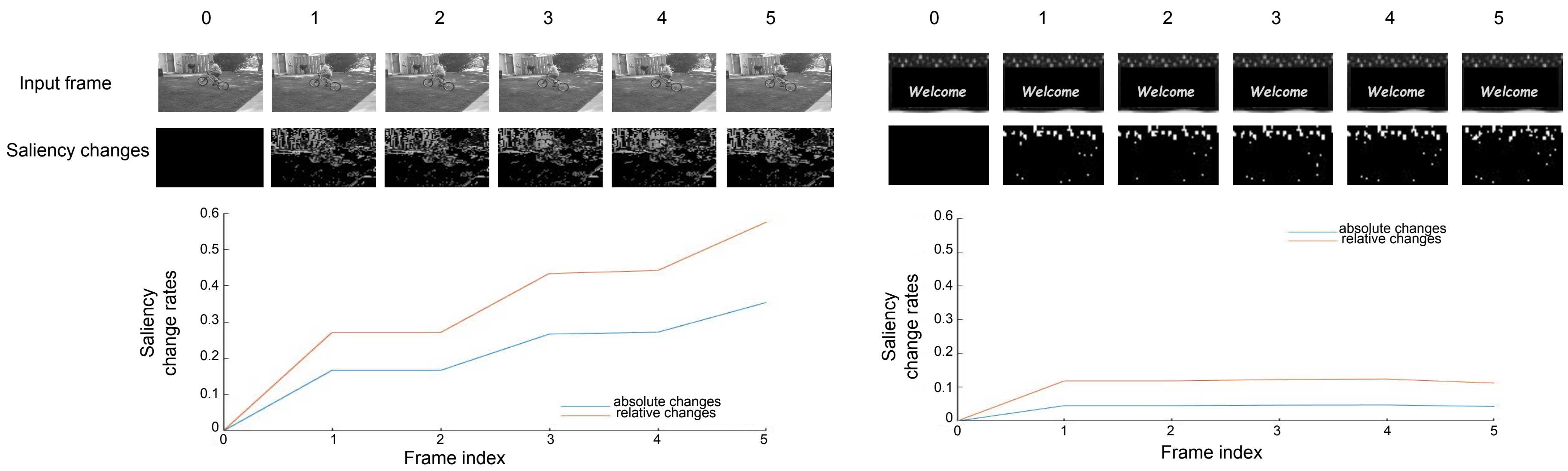
The major difference between the true jerkiness and the fake case is whether there is a smooth motion in the video !

Saliency change

- Notation
  - $I_i(x)$: intensity for macro block x on frame i
  - b_i : total number of macro blocks for frame i
  - m_i : number of masked macro blocks for frame i
  - T_{sc} : minimum value for a noticeable intensity change
- Saliency change
  - number of corresponding blocks that have noticeable differences between frame i and j
  - $sc_{i,j} = \sum_x (|I_i(x) - I_j(x)| > T_{sc})$
- Absolute and relative saliency change rate

$$abs\_rate = sc_{i,j}/b_i,$$
$$rel\_rate = sc_{i,j}/m_i.$$

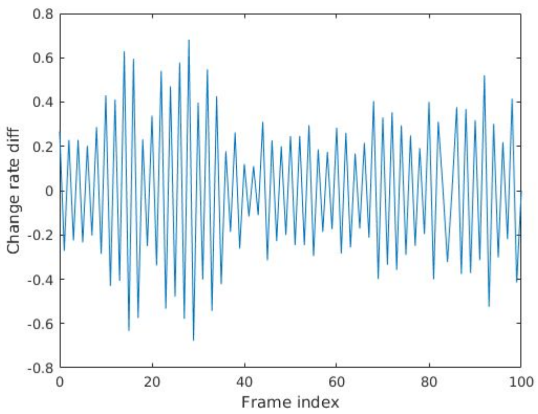
Saliency change rates



Our system stores previous frames in a buffer, and uses the median saliency change to decide whether there is a motion.

Jerkiness Feature Aggregation (in Post-processing)

- Three motion status for each video shot
  - no motion: $AVG_{\text{abs\_rate}} < MIN_{\text{abs\_rate}}$ && $AVG_{\text{rel\_rate}} < MIN_{\text{rel\_rate}}$
  - fast motion: $AVG_{\text{rel\_rate}} > MAX_{\text{rel\_rate}}$
  - smooth motion: otherwise.



Jerkiness artifacts exist if and only if there is certain cyclical pattern appearing in the profile of change rate diffs.

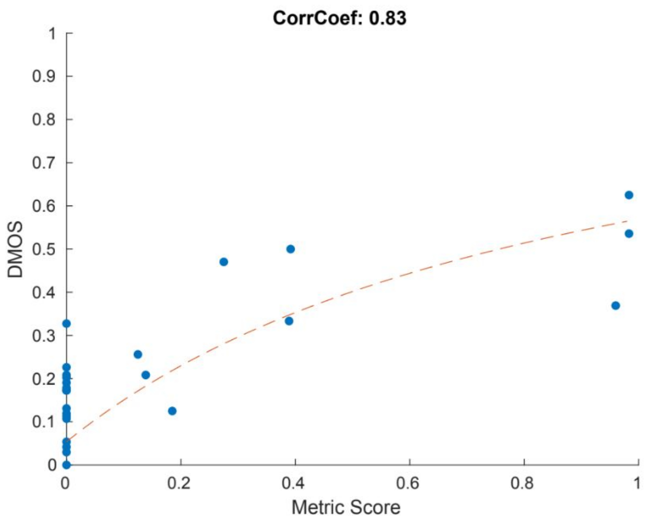
$$J = \max(0, 1 - \text{std}(dists[\cdot])) * ratio$$

distances between index with large change rate diffs

length of the section with cyclical pattern divided by the chunk length

Subjective Experiment

- 25 5s video clips are selected from 1,300 UGC videos, where some clips visually contain jerkiness artifacts



Fitted by a logistic model:

DMOS = 1 − 1/(c0 + exp(c1 · metric_score + c2)

where (c0, c1, c2) is (−9.32, 0.11, 2.34)

Final Quality Report

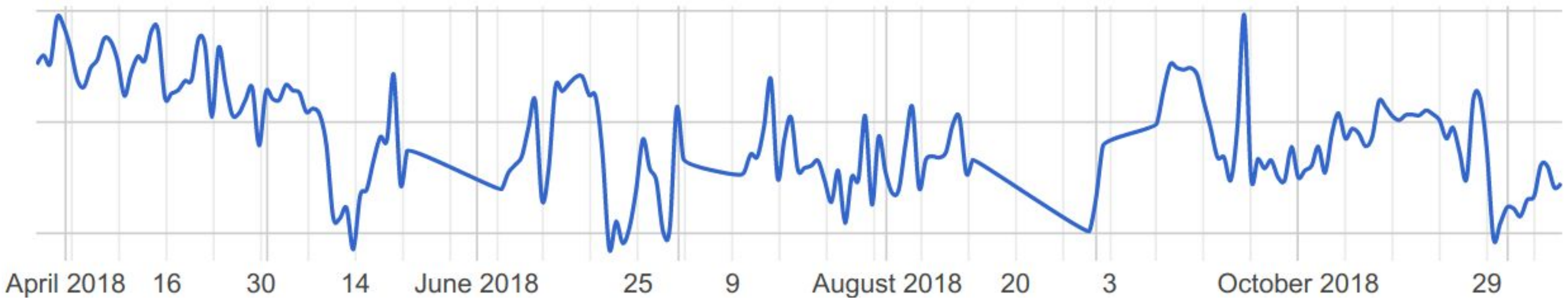
- Quality scores for all artifacts
  - e.g.

banding	noise	sleeq	jerkiness
0.6	0.2	0.1	0.5

  - relatively more flexible than just an overall score
- Provided adjustable quality bars for various use cases
  - e.g. banding = 0.6 may be OK for Lecture videos but BAD for Movies and Music videos

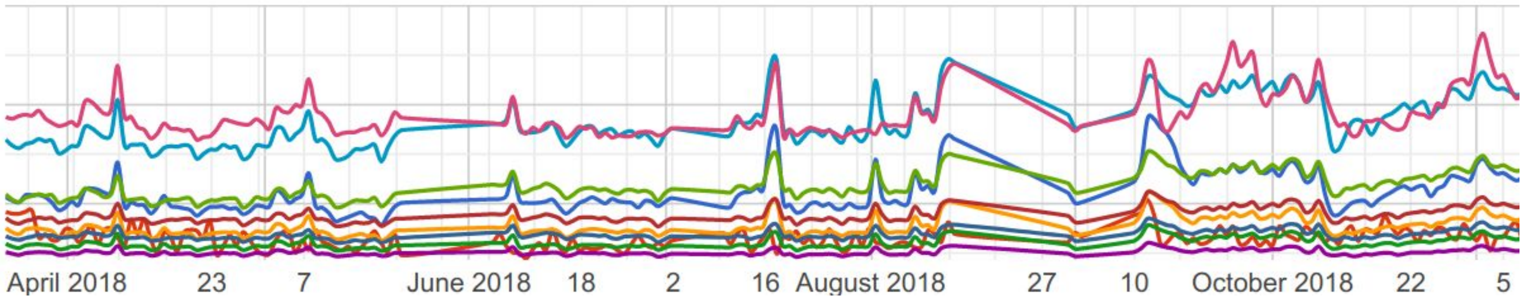
Quality Dashboard

Noise artifacts in uploaded videos



Quality Dashboard

Relative Banding: Transcoded_MOS - Original_MOS



Conclusion

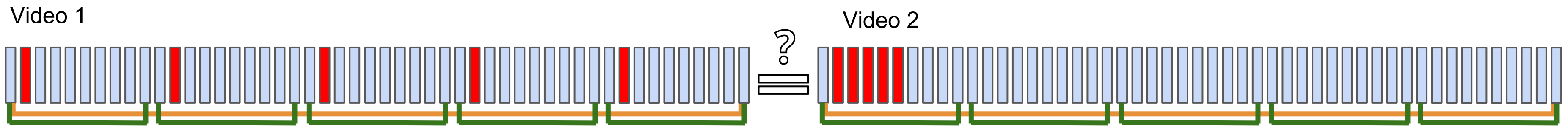
- In this talk, we
  - addressed challenges for UGC quality analysis
  - introduced a new framework for video quality analysis
  - introduced no reference features for quality evaluation
- In future, we will
  - keep exploring quality issues for UGC videos
  - release our UGC dataset

YouTube UGC Dataset

Thanks

Shortcomings of Traditional Framework

- Ambiguous frame quality aggregation.

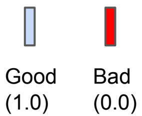


Traditional report:
Quality is 0.9 .

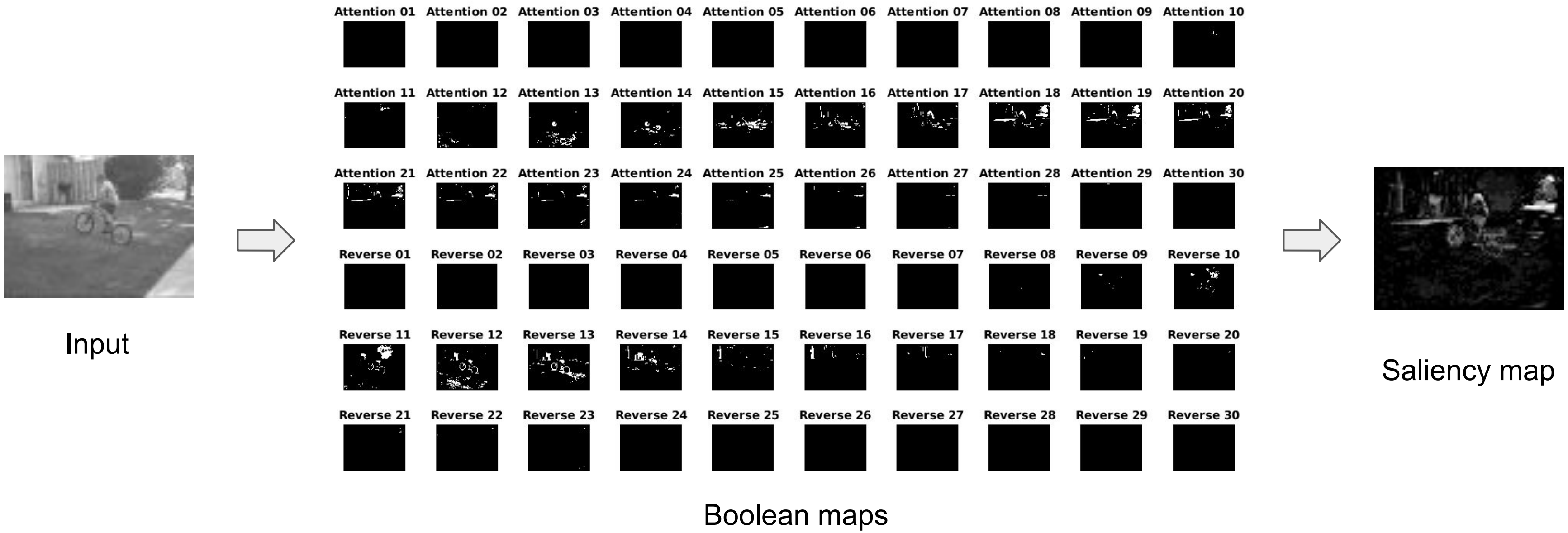
New report:
No chunk has bad quality.
Repeated bad quality frame detected.
Avg chunk quality is 0.9 .

Traditional report:
Quality is 0.9 .

New report:
20% chunks have bad quality.
Avg chunk quality is 0.9.
Worst chunk quality is 0.5, which starts from frame 0 to frame 9.

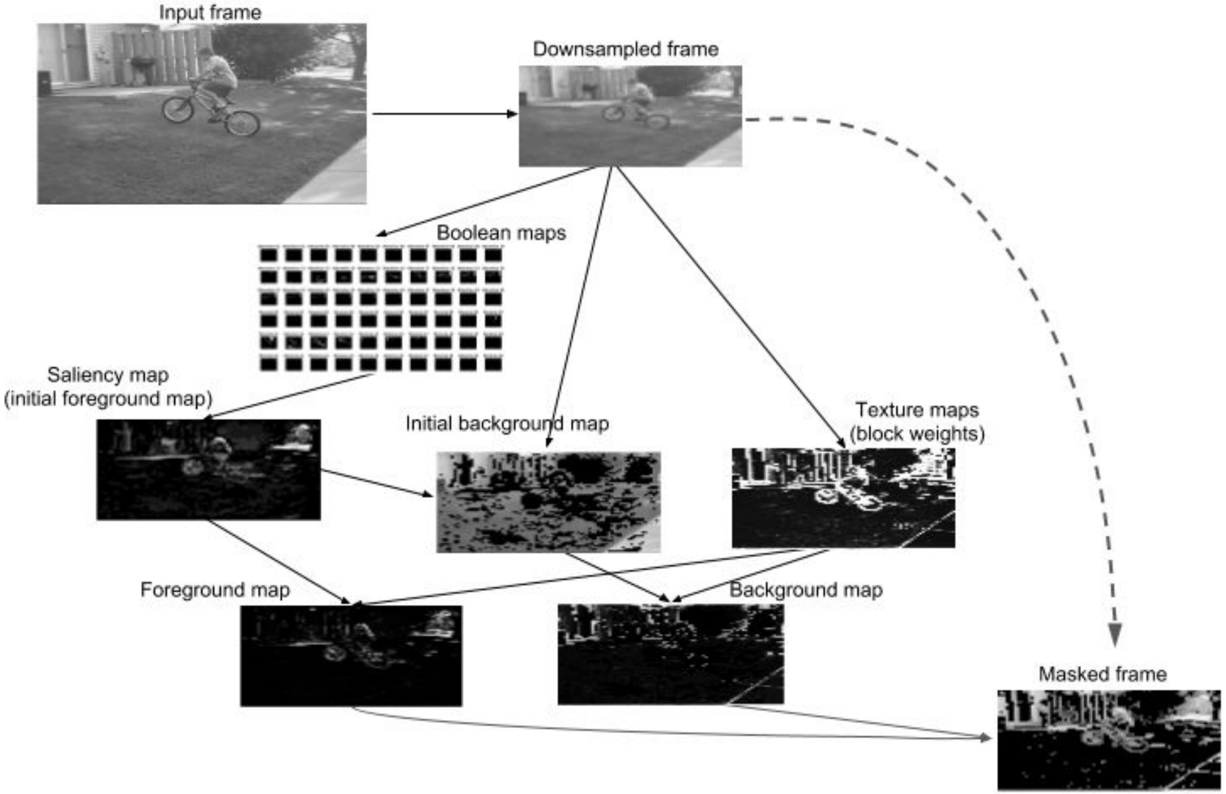


Boolean Map based Saliency (BMS)



Zhang, J. and Sclaroff, S., “Saliency detection: a boolean map approach,” ICCV 2013

Saliency Map Generation



Role of Saliency Map

- One application: to improve existing spatial quality metrics (e.g. SSIM)
- Suppose

$ssim_{\mathrm{b}}(i)$: SSIM score for block i

$w_{\mathrm{fore}}(i)$: foreground saliency weight for block i

$w_{\mathrm{back}}(i)$: background saliency weight for block i

then

$$Weighted\_SSIM = \begin{cases} 0, & \text{if } \sum_i w_{\mathrm{fore}}(i) = 0 \;\&\&\; \sum_i w_{\mathrm{back}}(i) = 0, \\ \sum_i (ssim_{\mathrm{b}}(i) * w_{\mathrm{back}}(i)) / \sum_i w_{\mathrm{back}}(i), & \text{if } \sum_i w_{\mathrm{fore}}(i) = 0, \\ \sum_i (ssim_{\mathrm{b}}(i) * w_{\mathrm{fore}}(i)) / \sum_i w_{\mathrm{fore}}(i), & \text{if } \sum_i w_{\mathrm{back}}(i) = 0, \\ \frac{\sum_i (ssim_{\mathrm{b}}(i) * w_{\mathrm{fore}}(i))}{2 \sum_i w_{\mathrm{fore}}(i)} + \frac{\sum_i (ssim_{\mathrm{b}}(i) * w_{\mathrm{back}}(i))}{2 \sum_i w_{\mathrm{back}}(i)}, & \text{otherwise.} \end{cases}$$

Experiments

- Weighted SSIM
  - LIVE Video Quality Assessment Database (10 original and 40 distorted videos).