Advances In Quality Assessment Of Video Streaming Systems: Algorithms, Methods, Tools

Yiannis Andreopoulos

Cosmin Stejerean

About Us

- Yiannis Andreopoulos is CTO of iSIZE Ltd., as well as Professor at University College London, UK. His expertise is in signal processing, machine learning and video streaming systems.

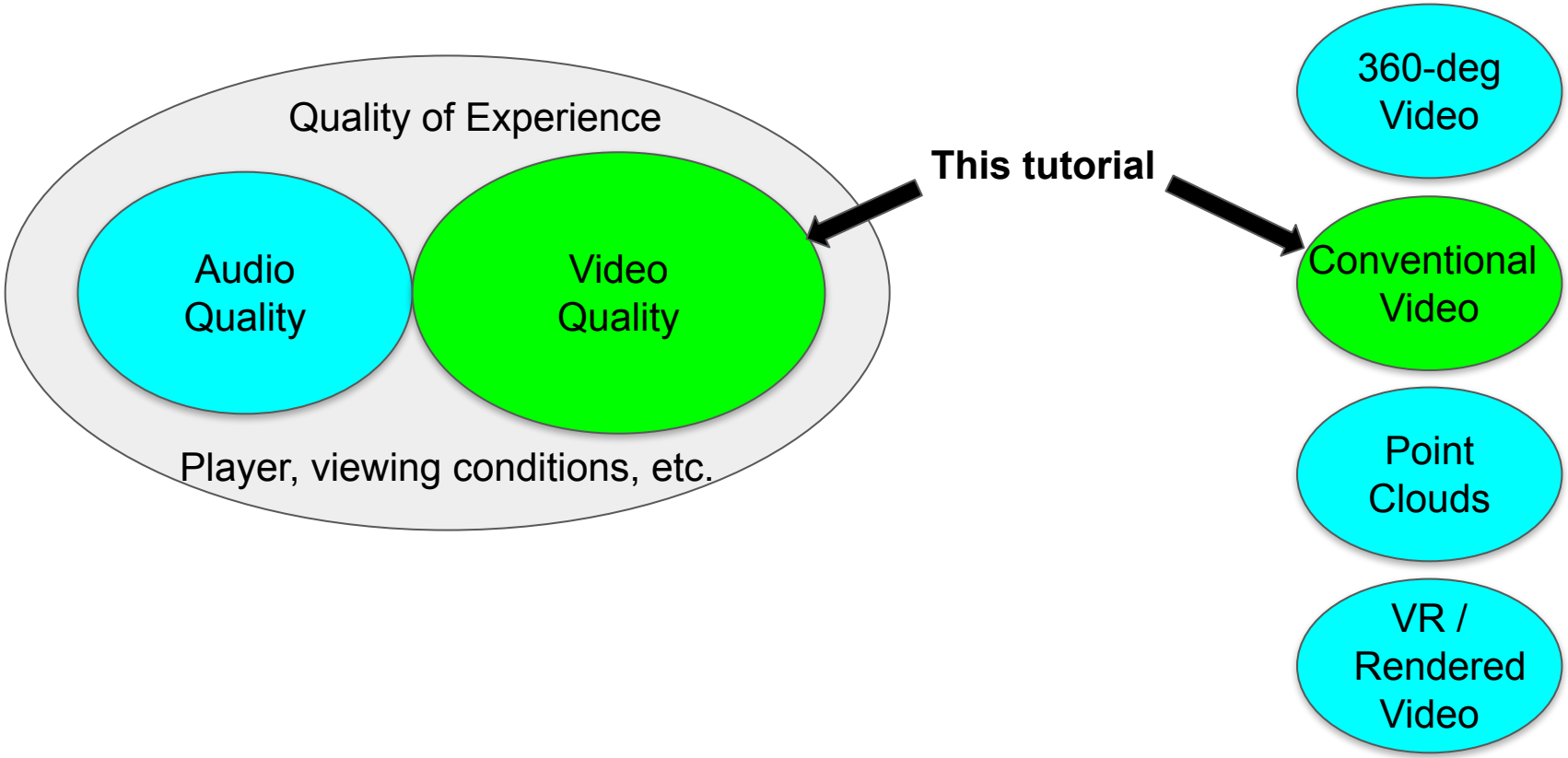
- Cosmin Stejerean is an engineer working on optimizing the quality of video at scale. He is a vice chair of the No Reference Metrics (NORM) project of the Video Quality Experts Group (VQEG). Cosmin’s research interests are in improving video quality assessment methods.

All material and opinions of this tutorial are of the presenters and do not represent iSIZE or Meta.
No endorsement is made (or implied) for any method or video material presented in this tutorial.

Tutorial Outline

- Video streaming, distortion, perception, quality assessment
- Quality metrics and subjective quality assessment
- Example use cases at scale
- Tools
- Future of quality assessment

Setting the Stage



Video Streaming: Real World Examples

Source: iSIZE, original video from XIPH.org

Video Streaming: Real World Examples

Source: iSIZE, original video from the YouTube UGC dataset media.withyoutube.com

Video Streaming: Real World Examples

Source: iSIZE, original video from the XIPH.org (AV2CTC)

Video Streaming: Real World Examples

Source: iSIZE, original video from the YouTube UGC dataset media.withyoutube.com

Video Streaming: Some Observations

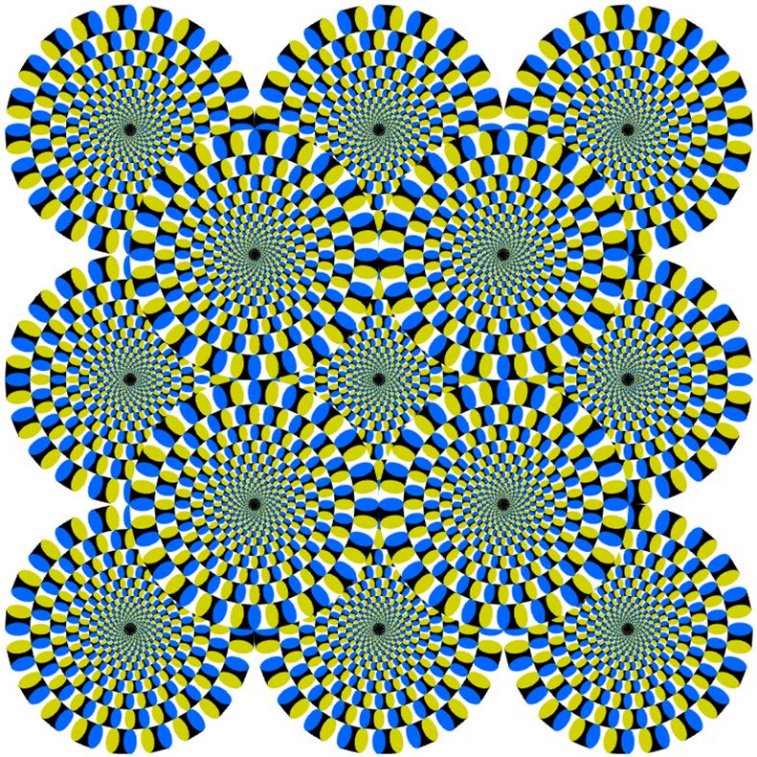
- Spatial and temporal masking, viewing conditions matter!
- What are the main sources of distortion?
- What is the video data manifold?
- What is distortion and what is artistic effect?

Spatial Aliasing Examples and Peripheral Vision

Notice:
Some of the pictures of the following 2 slides can cause dizziness or in some very rare cases might possibly cause epileptic seizures. The latter happens when the brain can't handle the conflicting information from your two eyes.

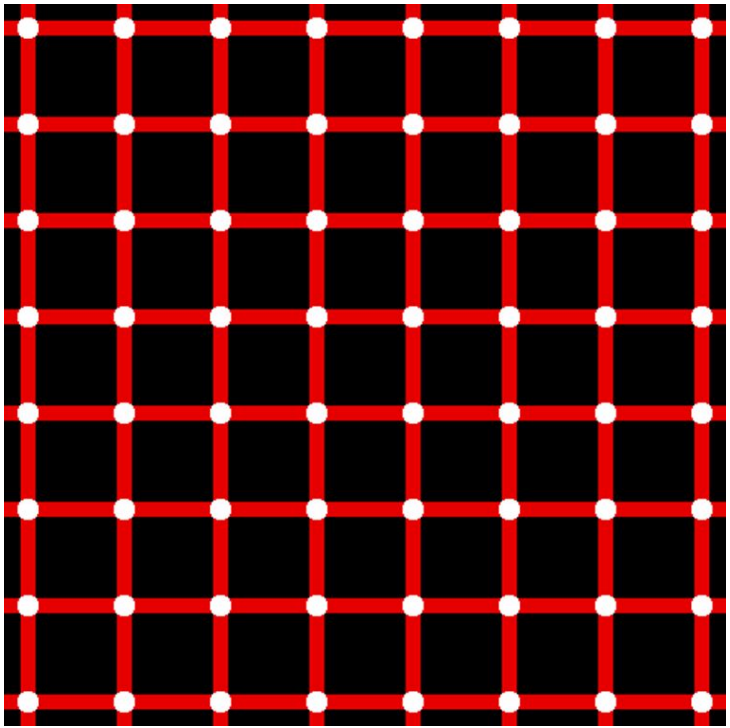
If you start feeling unwell when viewing the slides, cover one eye with your hand immediately and then look away from the screen. Do *not* close your eyes because that can make the attack worse.

Spatial Aliasing Examples and Peripheral Vision



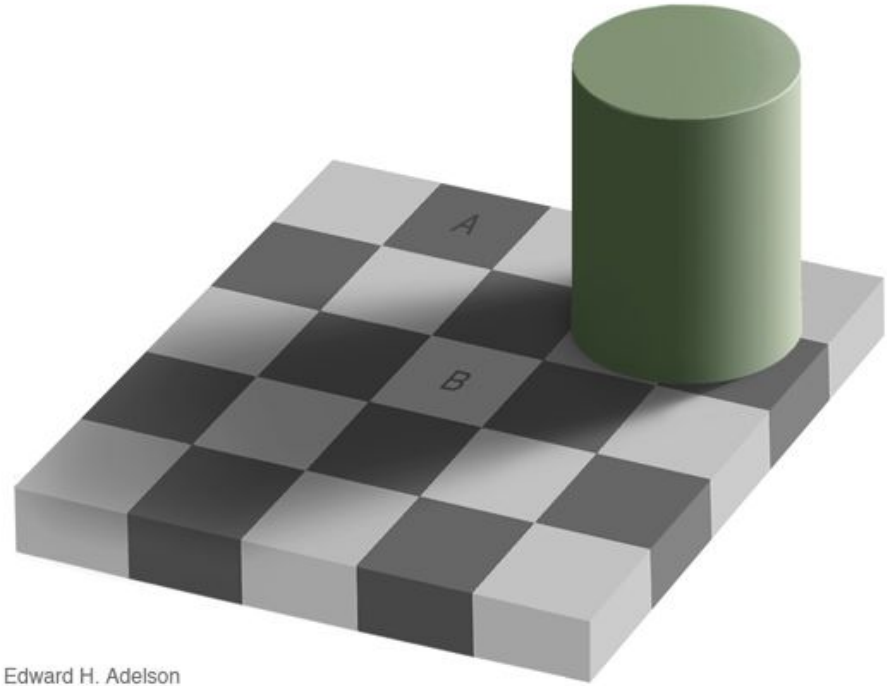
https://www.ritsumei.ac.jp/~akitaoka/index-e.html

Spatial Aliasing Examples and Peripheral Vision



Schrauf M, Lingelbach B & Wist, “The scintillating grid illusion,” Vision Res 37:1033–1038 , 1997

Contrast Masking (Cornsweet Illusion & Past Experience)



Edward H. Adelson

https://en.wikipedia.org/wiki/Checker_shadow_illusion

Main Sources of Distortion

- Sampling and quantization
- DPCM in video coding
- Different prediction modes
- Rate control, ABR ladder adaptation
- Advanced tools (warping, AI-based encoding)

Main Sources of Distortion

Original Lena Image = 262144 bytes

Main Sources of Distortion

Result of Lena Compressed 16 times to 16384 bytes

Main Sources of Distortion

Result of Lena Compressed 32 times to 8192 bytes

Main Sources of Distortion

Result of Lena Compressed 64 times to 4096 bytes

Main Sources of Distortion

Result of Lena Compressed 128 times to 2048 bytes

2048 bytes =
1024 words =
1K words
This picture
is one thousand words!

Main Sources of Distortion

Result of Lena Compressed 256 times to 1024 bytes

Main Sources of Distortion

Result of Lena Compressed 512 times to 512 bytes

Main Sources of Distortion

Where is the loss coming from?

1) Sampling
2) Quantization

Original sampling:
512 rows × 512 columns

Main Sources of Distortion

Where is the loss coming from?

Subsampled by 4:
128 rows × 128 columns

Main Sources of Distortion

Where is the loss coming from?

Original quantization:
8 bits per pixel

Main Sources of Distortion

Where is the loss coming from?

Quantized to 4 bits per pixel

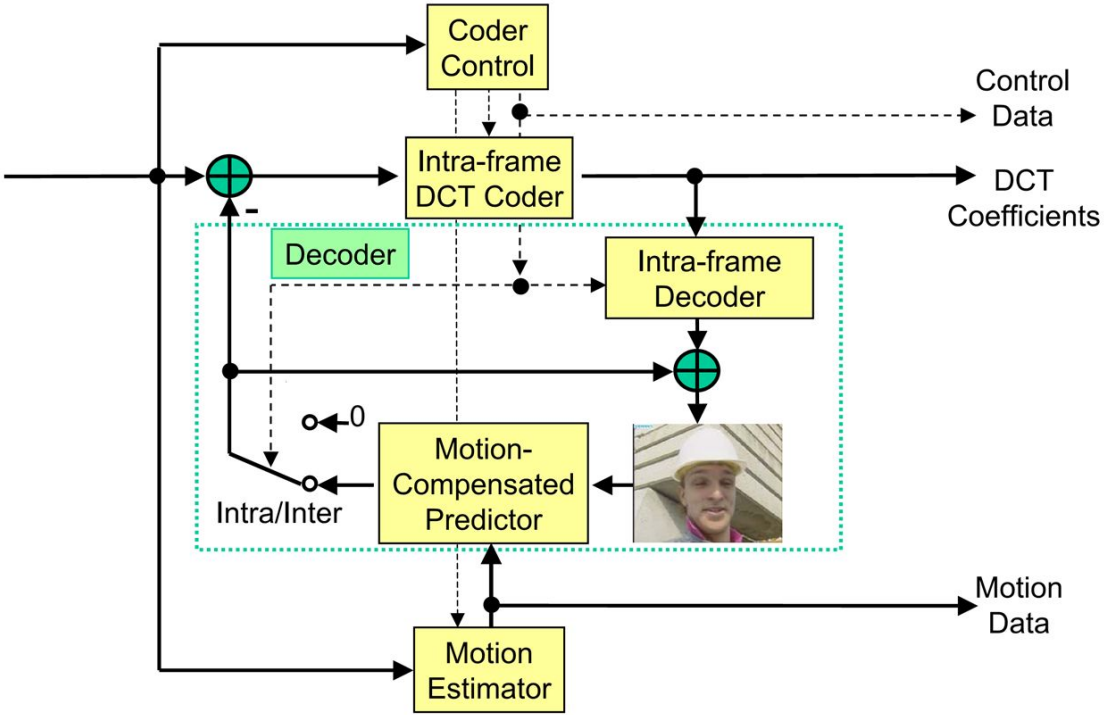
Main Sources of Distortion

Where is the loss coming from?

Quantized to 2 bits per pixel

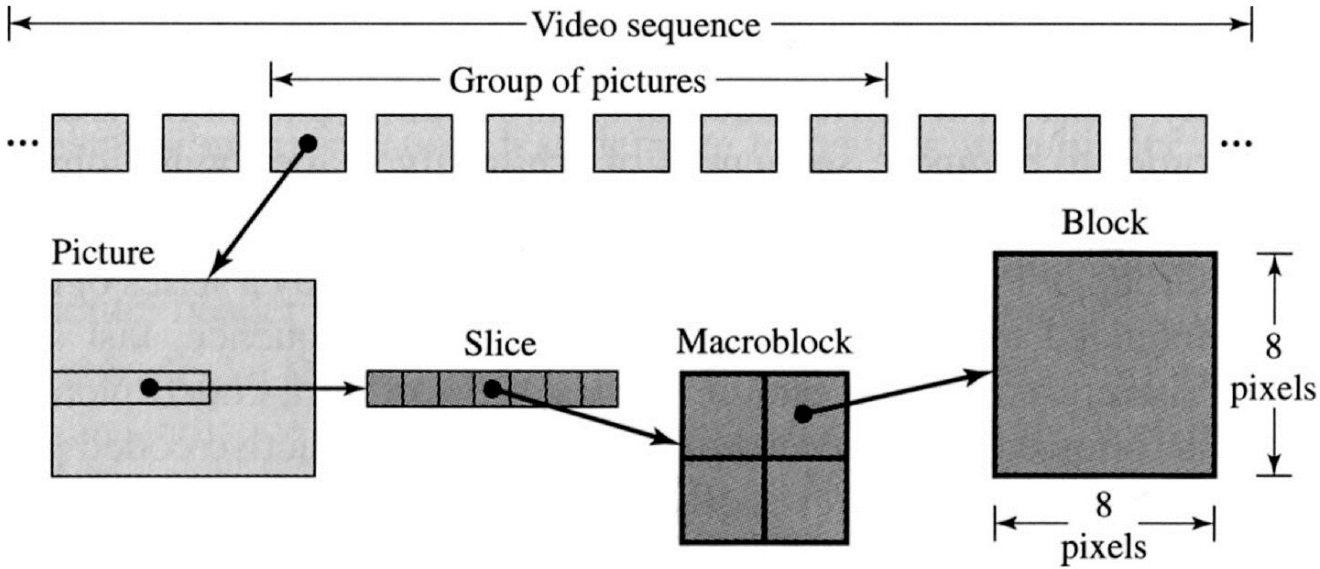
Main Sources of Distortion

- DPCM, a.k.a., closed-loop prediction



Main Sources of Distortion

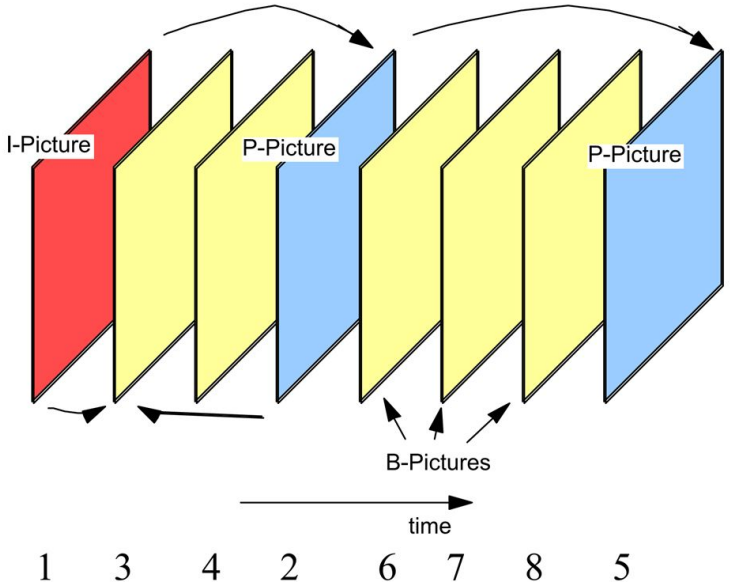
- Different prediction modes per picture, slice, macroblock and block



Main Sources of Distortion

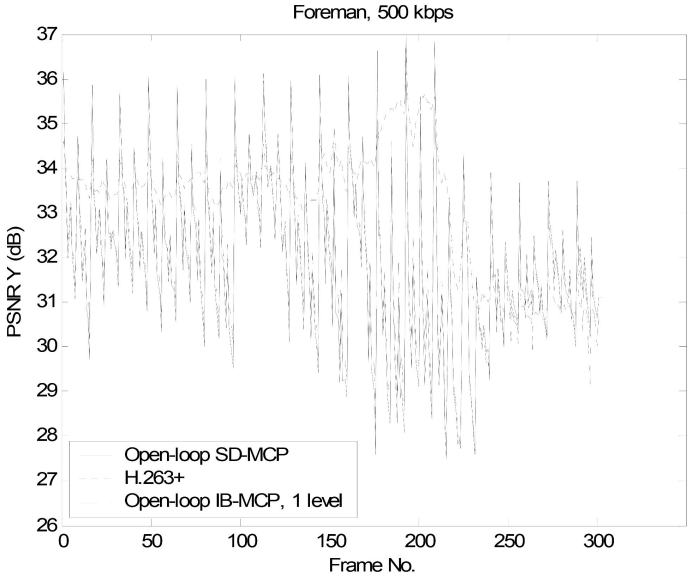
- At a high-level, known as picture types

"Group of Pictures" = “GOP“, GOP structure is very flexible



Main Sources of Distortion: Temporal Flickering

- Rate control

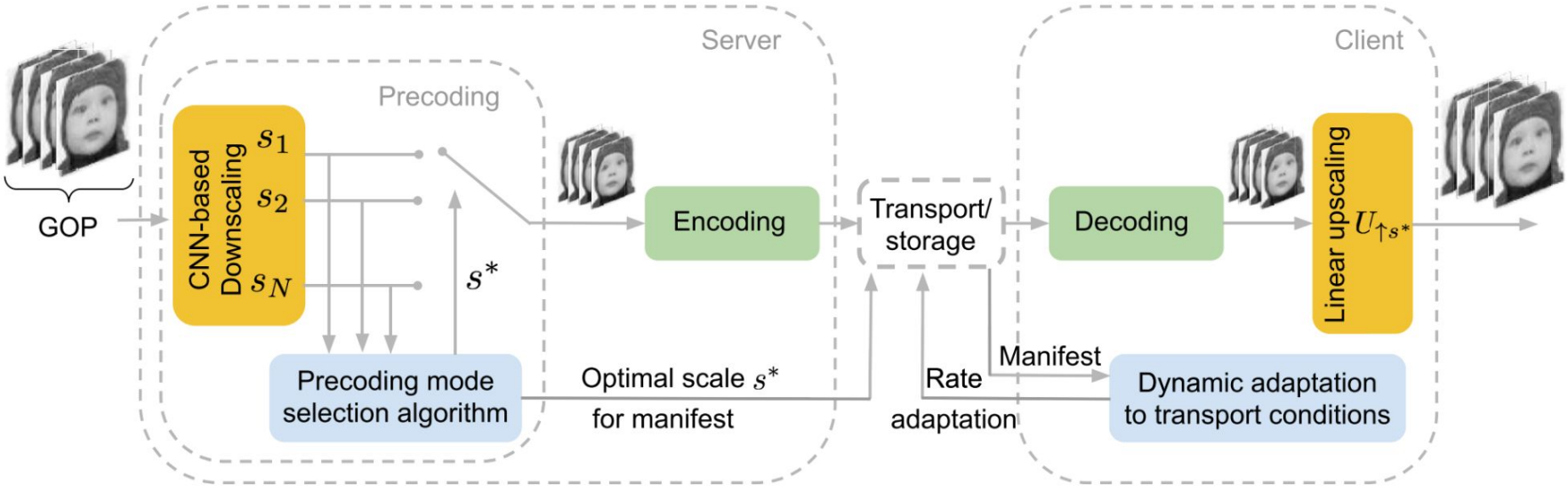


[Y. Andreopoulos, PhD thesis]

```
ffmpeg -i foreman.y4m -c:v libx264 -qp 33 -g 5
-i_qfactor 10 foreman_b.mp4
```

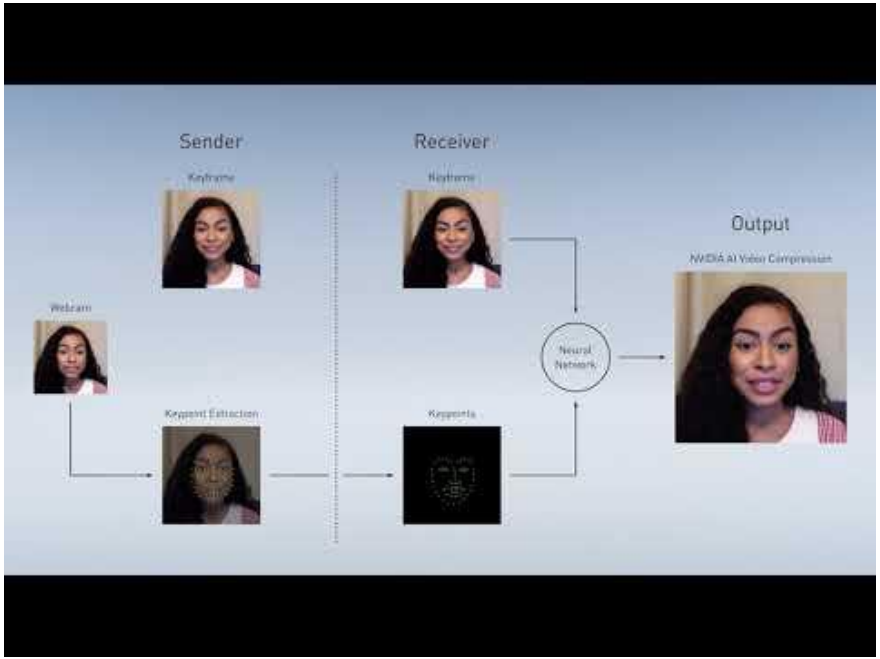
Main Sources of Distortion

- Shot-based encoding, ABR ladder adaptation



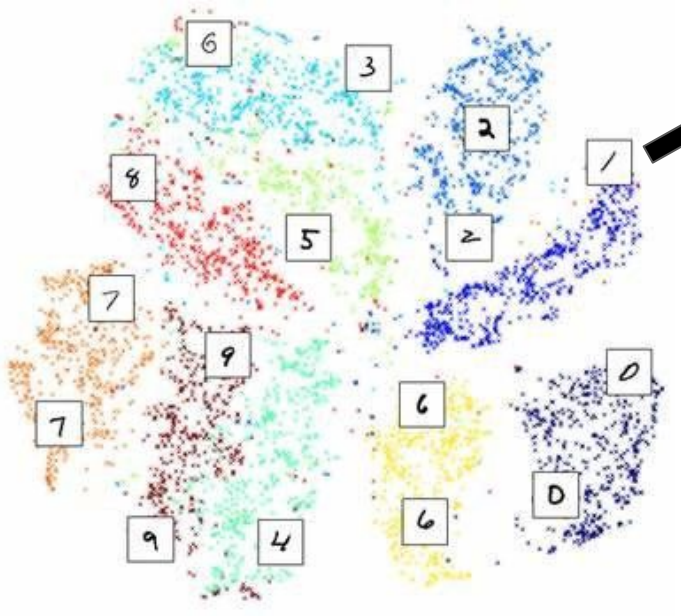
Main Sources of Distortion

- Advanced encoding tools (keypoint-based rendering, warping, AI-based encoding)



What is the Video Data Manifold?

- For the case of video, this can be extremely complex…
- What about animation, gaming, artistic effects?
- What is distortion and what is artistic effect?

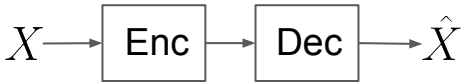


$P(X)$: the distribution of natural video R.V. X

https://towardsdatascience.com/manifolds-in-data-science-a-brief-overview-2e9dde9437e5

What is the Video Data Manifold?

- Perceptual video quality = the degree to which a video looks like a natural video
  → Human mean opinion scores
  → No-reference metric
  → Real/fake tests
  → Example divergence measures: total variation, Wasserstein distance, *f*-divergence, etc.



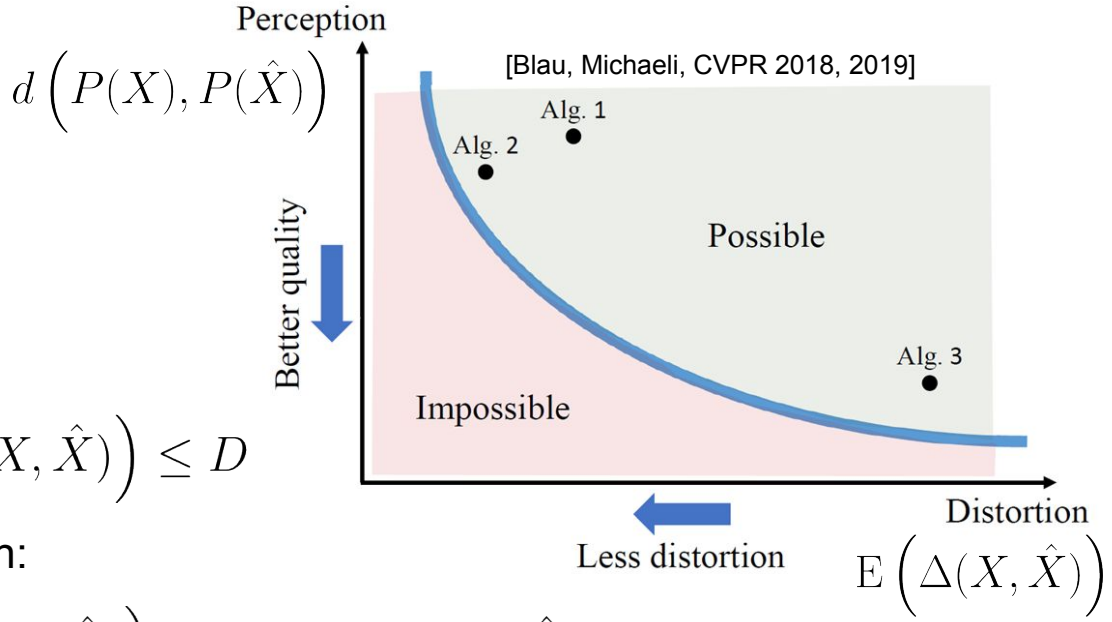
Rate: $I(X, \hat{X})$

Distortion: $\mathrm{E}\left(\Delta(X, \hat{X})\right)$

Perception: $d\left(P(X), P(\hat{X})\right)$

Source VVC@9.2kbps iSIZE BitGen@2.5kbps

Rate-Perception-Distortion Trade-off



Traditional rate-distortion optimization:

$$R(D) = \min\{I(X, \hat{X})\} \text{ s.t. } \mathrm{E}\left(\Delta(X, \hat{X})\right) \leq D$$

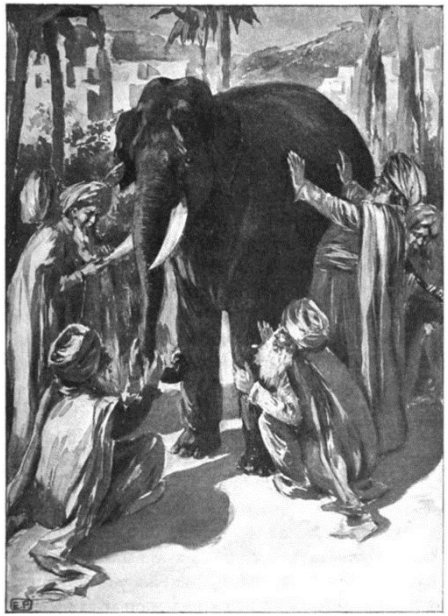
Rate-distortion-perception optimization:

$$R(D) = \min\{I(X, \hat{X})\} \text{ s.t. } \mathrm{E}\left(\Delta(X, \hat{X})\right) \leq D, d(P(X), P(\hat{X}) \leq Q$$

Blau shows that if $d\left(P(X), P(\hat{X})\right)$ is convex for $P(\hat{X})$ then the perception-distortion function is monotonically non-increasing and convex

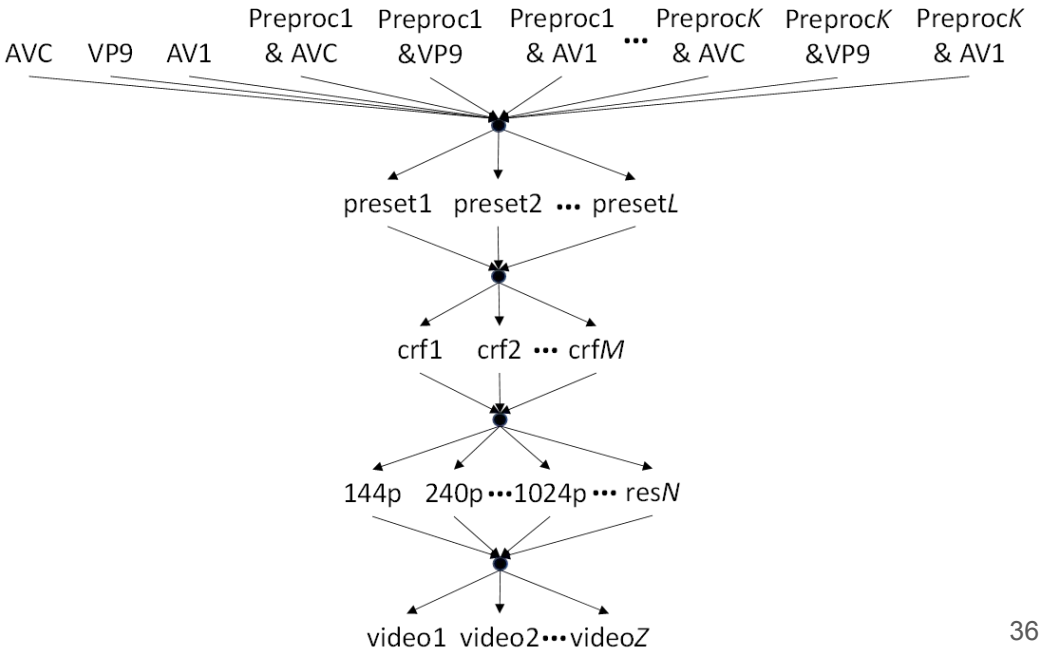
The Three Challenges of Video Quality Assessment

1. Objective metrics (and humans) are myopic



https://en.wikipedia.org/wiki/Blind_men_and_an_elephant

2. The exploration space can surpass 1m tests for a 100-video library



The Three Challenges of Video Quality Assessment

3. Video streaming algorithms are now increasingly optimized for perceptual quality metrics instead of signal distortion

Source | VVC@147kbps | iSIZE BitGen@31kbps

Tutorial Outline

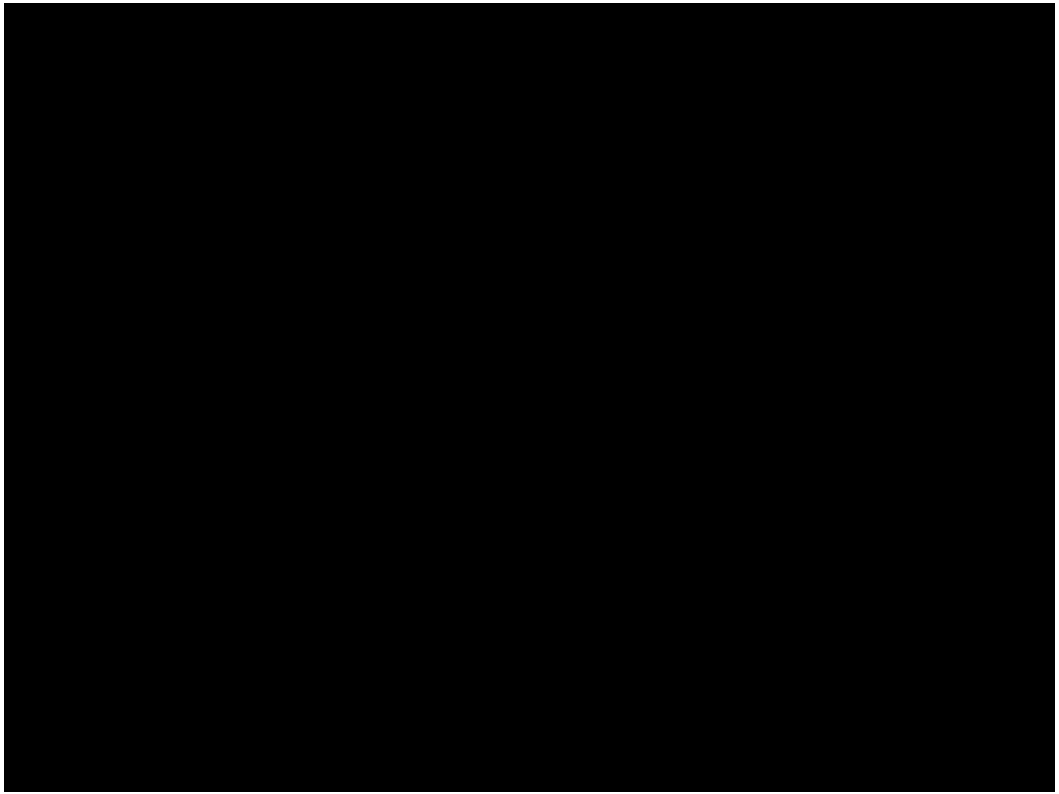
- Video streaming, distortion, perception, quality assessment
- Quality metrics and subjective quality assessment
- Example use cases at scale
- Tools
- Future of quality assessment

Subjective Quality Assessment

- All key information is available in ITU-T P.913 or BT.500 standards
- Requires careful tuning of room conditions, display device, distance from the screen, scores collection and post-processing
- The player, clips duration and fps needs to be aligned
- The participants must be screened for their eyesight and color blindness

Subjective Quality Assessment: ACR or DCR?

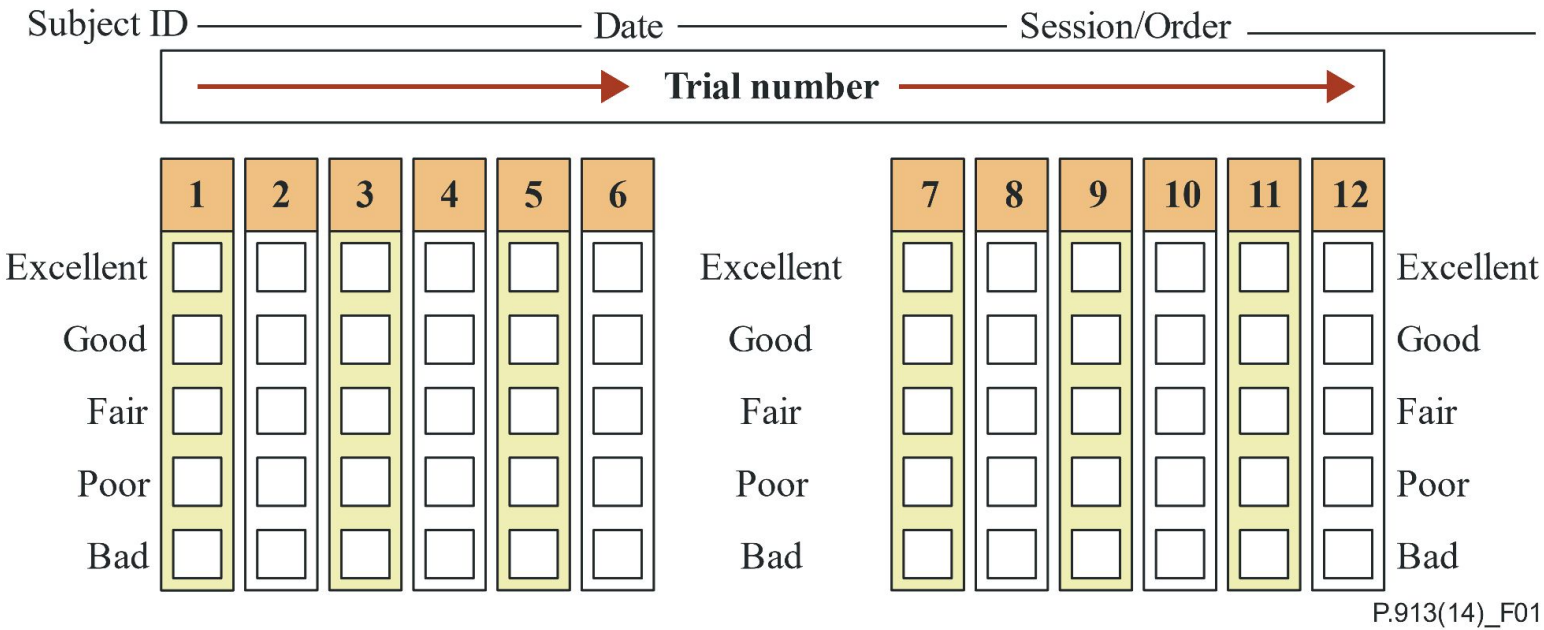
- DCR



Subjective Quality Assessment: ACR or DCR?

- ACR

Subjective Quality Assessment: Example Scoresheet



[ITU-T P.913, online at:https://www.itu.int/rec/T-REC-P.913-202106-I/en]

Subjective Quality Assessment: Other Methods + Parameters

- Q1: Which method is the most accurate, fastest and easiest?
- Q2: Which rating scale to use? 5-point/10-point/11-point…

Answer: ACR-HR 5-point scale

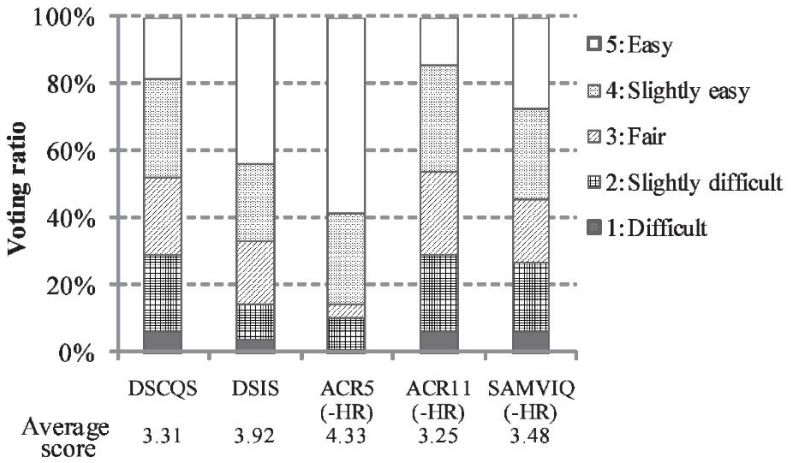
MCI_{norm} for each method

	DSCQS	DSIS	ACR5	ACR5-HR	ACR11	ACR11-HR	SAMVIQ	SAMVIQ-HR
MCInorm	0.09	0.07	0.07	0.09	0.08	0.10	0.07	0.08

Total assessment time (minutes)

Method	Average	Max.	Min.	Std.
DSCQS	41	45	37	2
DSIS	20	25	16	2
ACR5 (-HR)	12	18	11	1
ACR11 (-HR)	14	20	11	3
SAMVIQ (-HR)	29	38	23	5

[Tominaga, QoMEX 2010]



Ease of evaluation for each method

Post-processing of Scores

- Q1: What is the best post-processing method?
- Q2: How can we handle outlier removal, subject bias and inconsistency?

  Answers: Netflix SUREAL, online at: https://github.com/Netflix/sureal

Model: $X_{e,s} = x_e + B_{e,s} + A_{e,s}$

$X_{e,s}$ the R.V. of raw score of video e from subject s

x_e the quality of video e by an average user

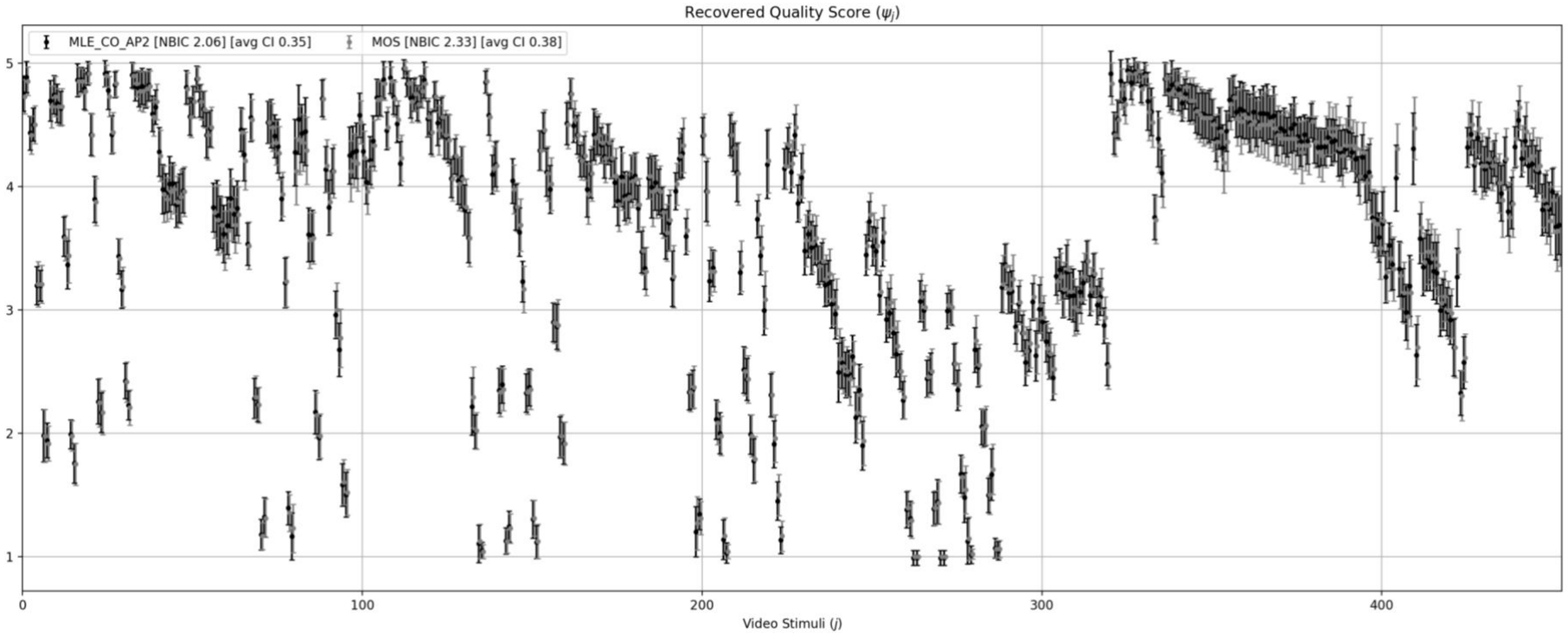
$B_{e,s} \sim N(b_s, v_s^2)$ the factor of subject s (i.i.d.)

$A_{e,s} \sim N(0, a_{c(e)}^2)$ the factor of content e (i.i.d.)

and the solution for x_e is obtained by maximum likelihood estimation

→ **No removal of outlier scores or subjects, only MLE-based adjustments!**

Post-processing of Scores



[Netflix SUREAL, online at: https://github.com/Netflix/sureal]

Quality Metrics: Reference-based

- PSNR, SSIM
- VMAF
- AVQT
- AI-based

Quality Metrics: Peak Signal to Noise Ratio

- $\mathrm{PSNR} = 10 \log_{10} \left(\frac{\mathrm{DYN\_RANGE}^2}{\mathrm{avg}\left((s_{i,j} - c_{i,j})^2\right)} \right)$, for two image arrays **S**, **C**

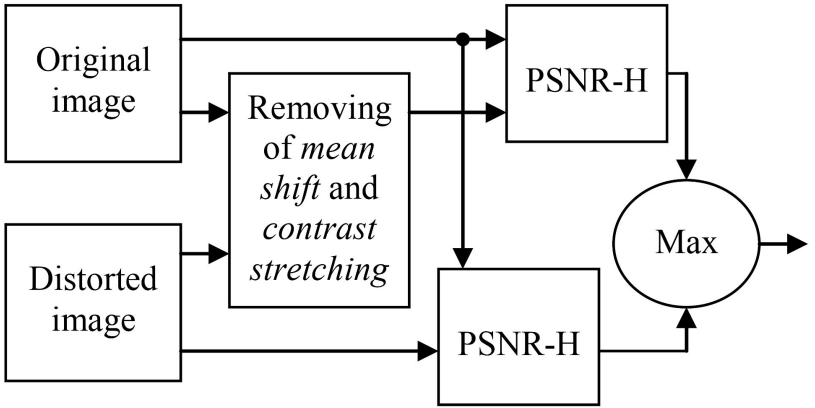
Has been extended to contrast perception and visual masking (PSNR-HVS/HVSM) [Ponomarenko, Carli, et al.], i.e.,

PSNR-H=PSNR with MSE the weighted MSE with the normalized JPEG 8x8 quantization table value

(+) fast, well-understood

(–) not accurate vs. P.910 MOS

(–) not normalized to 0-100 scale



Quality Metrics: Structural Similarity Index Metric

- $\mathrm{SSIM} = l(\mathbf{X},\mathbf{Y})c(\mathbf{X},\mathbf{Y})s(\mathbf{X},\mathbf{Y})$

  with $l(\mathbf{X},\mathbf{Y}) = \dfrac{2\mu_{\mathbf{X}}\mu_{\mathbf{Y}} + c_1}{\mu_{\mathbf{X}}^2 + \mu_{\mathbf{Y}}^2 + c_1}$, $c(\mathbf{X},\mathbf{Y}) = \dfrac{2\sigma_{\mathbf{X}}\sigma_{\mathbf{Y}} + c_2}{\sigma_{\mathbf{X}}^2 + \sigma_{\mathbf{Y}}^2 + c_2}$, $s(\mathbf{X},\mathbf{Y}) = \dfrac{2\sigma_{\mathbf{XY}} + c_3}{\sigma_{\mathbf{X}}\sigma_{\mathbf{Y}} + c_3}$

  luminance contrast structure

  Has been extended to MS-SSIM and several other variations

  (+) fast, well-understood

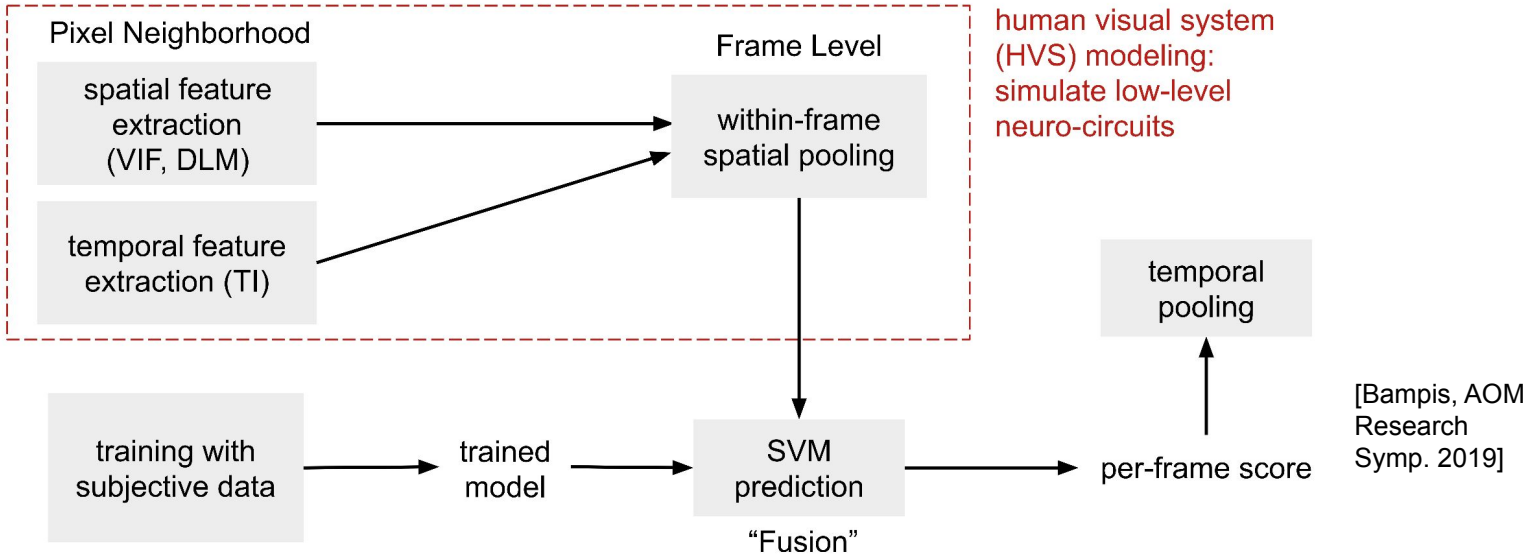
  (+) suited to resolution and viewing conditions

  (–) often not accurate enough

  (–) for video streaming, the SSIM scale is very narrow

Quality Metrics: Video Multimethod Assessment Fusion

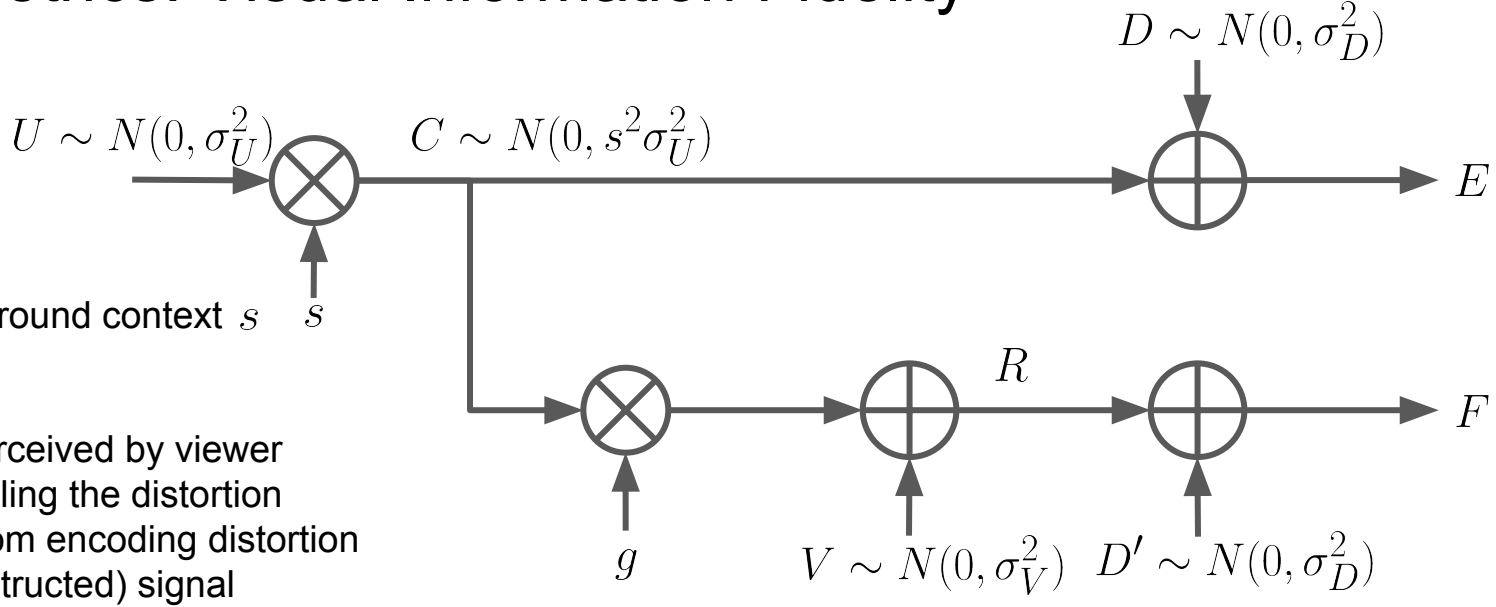
- VMAF = svr(DLM,VIF,motion)



(+) well supported by Netflix, has stood the test of time

(–) can be too slow to run at scale, no support yet for beyond-8bit content

Quality Metrics: Visual Information Fidelity



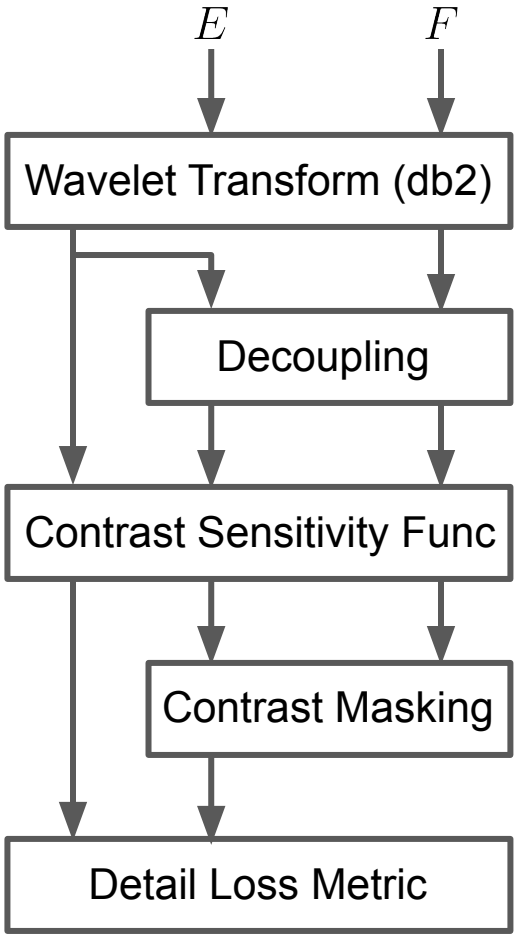
U : local variability around context s
C : source model
D : HVS model
E : source signal perceived by viewer
g : gain term controlling the distortion
V : additive noise from encoding distortion
R : decoded (reconstructed) signal
D': HVS model for the decoded content
F : distorted signal perceived by viewer

$$g = \frac{\sigma_{CR}}{\sigma_C^2} \qquad \sigma_V^2 = \sigma_R^2 - g\sigma_{CR}$$

$$\text{per-scale VIF} = \frac{\sum_{\forall i} \log_2 \left(1 + \frac{g_i^2 s_i^2 \sigma_U^2}{\sigma_{V_i}^2 + \sigma_D^2}\right)}{\sum_{\forall i} \log_2 \left(1 + \frac{s_i^2 \sigma_U^2}{\sigma_D^2}\right)}$$

Quality Metrics: Detail Loss Metric

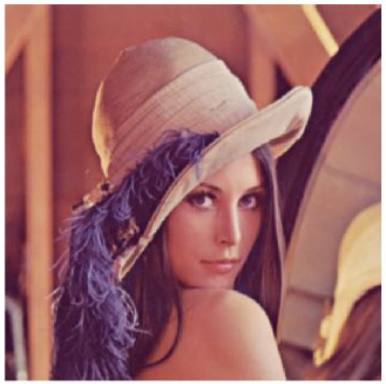
- Wavelet decomposition with db2 filters and gain O between F and E is calculated per subband and per coefficient
- Contrast sensitivity function: $H(\omega) = (a + b\omega)\exp(-c\omega)$ (adjusted to picture height, viewing distance & cpd)
- Contrast masking adjusts to psychovisual experiments of masking effects near similar neighboring spatial freq.
- The coefficients in after CSF and CM are then Minkowski-pooled with power 3, and summed within the center region of each subband and scale



Example 1: Response to Image Blur and Speckle Noise

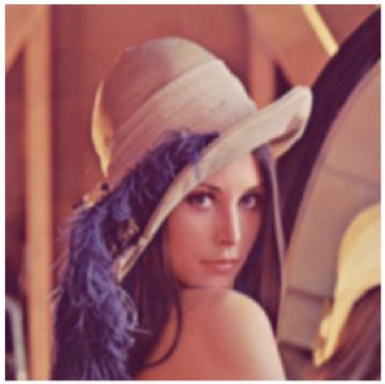
Original

Gaussian Blur 1



psnr=38.4, ssim=0.98, vif=0.63

Gaussian Blur 2



psnr=35.1, ssim=0.96, vif=0.33

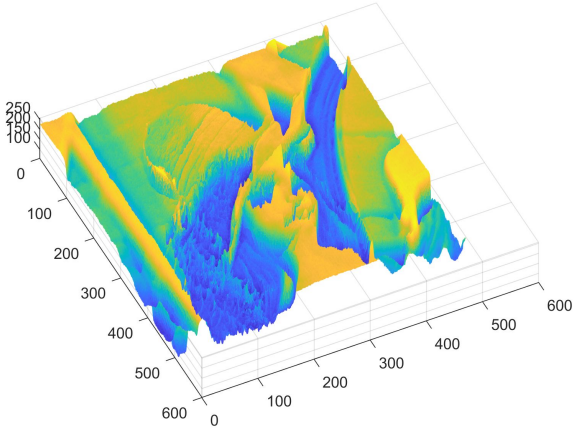
Speckle Noise

psnr=38.1, ssim=0.89, vif=0.23

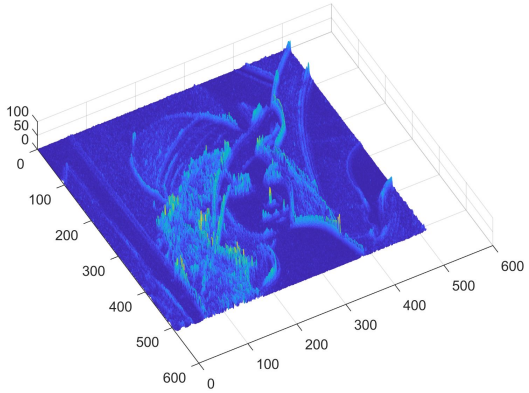
```
z=imread('lena_color.tiff');
H1=fspecial('gaussian',11,1); H2=fspecial('gaussian',21,2);
zlow1=imfilter(z,H1,'symmetric','same'); zlow2=imfilter(z,H2,'symmetric','same');
N=100; pos=floor(rand(1,N)*512)+1; zcorr=z; zcorr(pos(1:N),pos(1:N),1)=0;
ssimval_low1=ssim(z,zlow1); ssimval_low2=ssim(z,zlow2);
vif_low1 = vifvec(z(:,:,1),zlow1(:,:,1)); vif_low2 = vifvec(z(:,:,1),zlow2(:,:,1));
psnr_low1=10*log10(255^2/(mean(mean((z(:,:,1)-zlow1(:,:,1)).^2))));
psnr_low2=10*log10(255^2/(mean(mean((z(:,:,1)-zlow2(:,:,1)).^2))));
ssimval_corr=ssim(z,zcorr); vif_corr = vifvec(z(:,:,1),zcorr(:,:,1));
psnr_corr=10*log10(255^2/(mean(mean((z(:,:,1)-zcorr(:,:,1)).^2))));
% Note: vif_vec code from: https://github.com/sattarab/image-quality-tools/tree/master/metrix_mux/metrix/vif
```

Example 1: Response to Image Blur and Speckle Noise

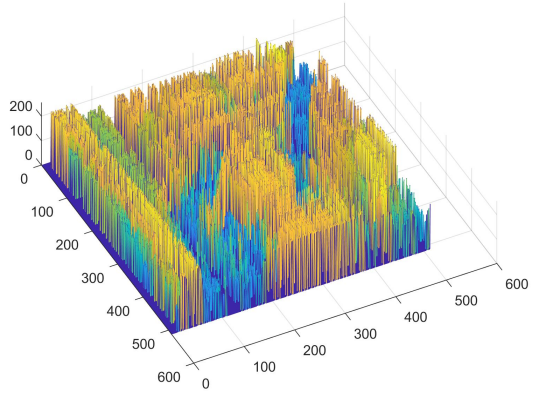
Original Image Surface



Gaussian Blur Error Surface

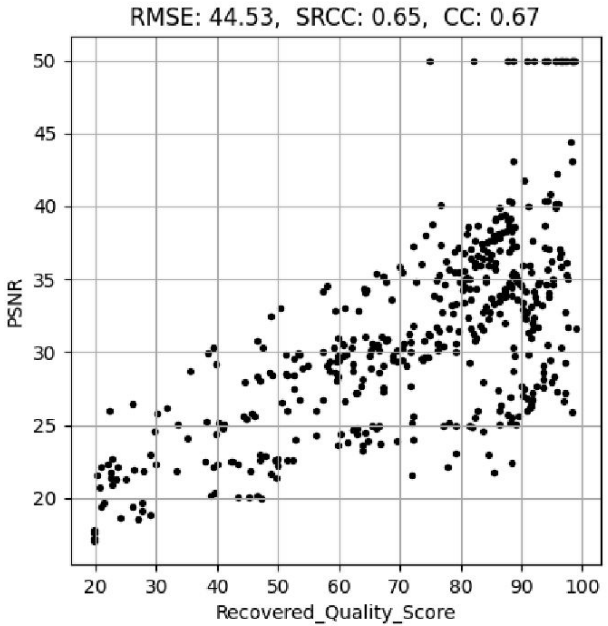


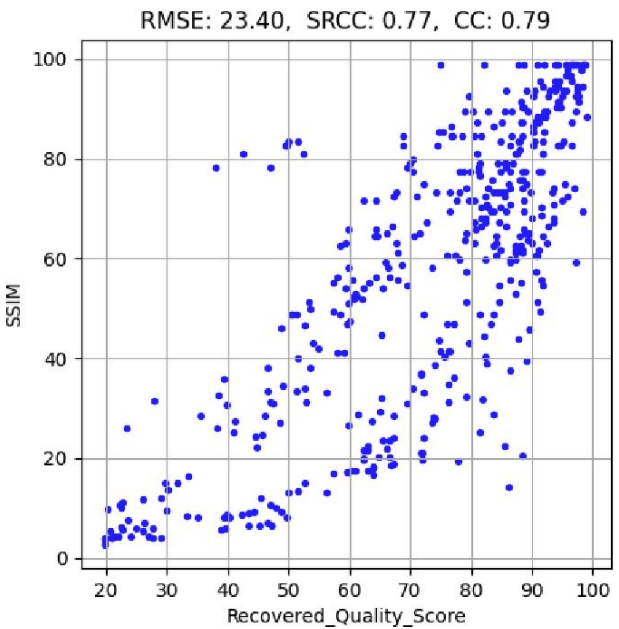
Speckle Noise Error Surface

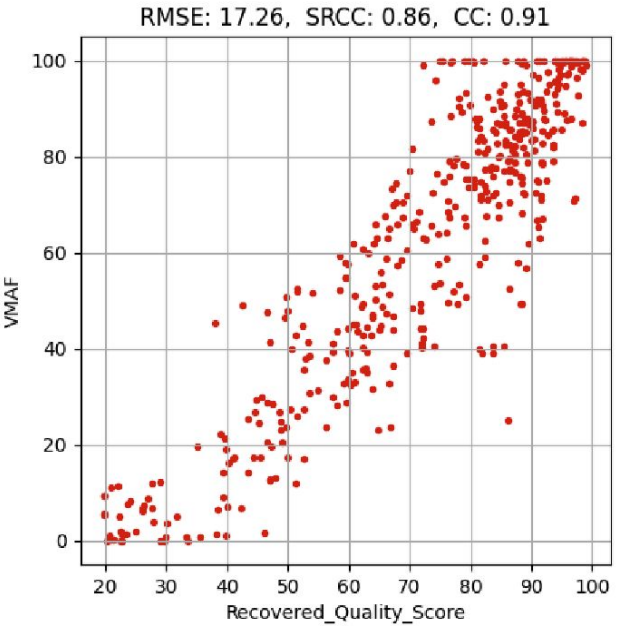



```
figure; mesh(z(:,:,1));
figure; mesh(transpose(round(sqrt((double(z(:,:,1))-double(zlow(:,:,1))).^2))));
figure; mesh(transpose(round(sqrt((double(z(:,:,1))-double(zcorr(:,:,1))).^2))));
```

Example 2: Fit to P.910 MOS







Quality Metrics: Advanced Video Quality Metric

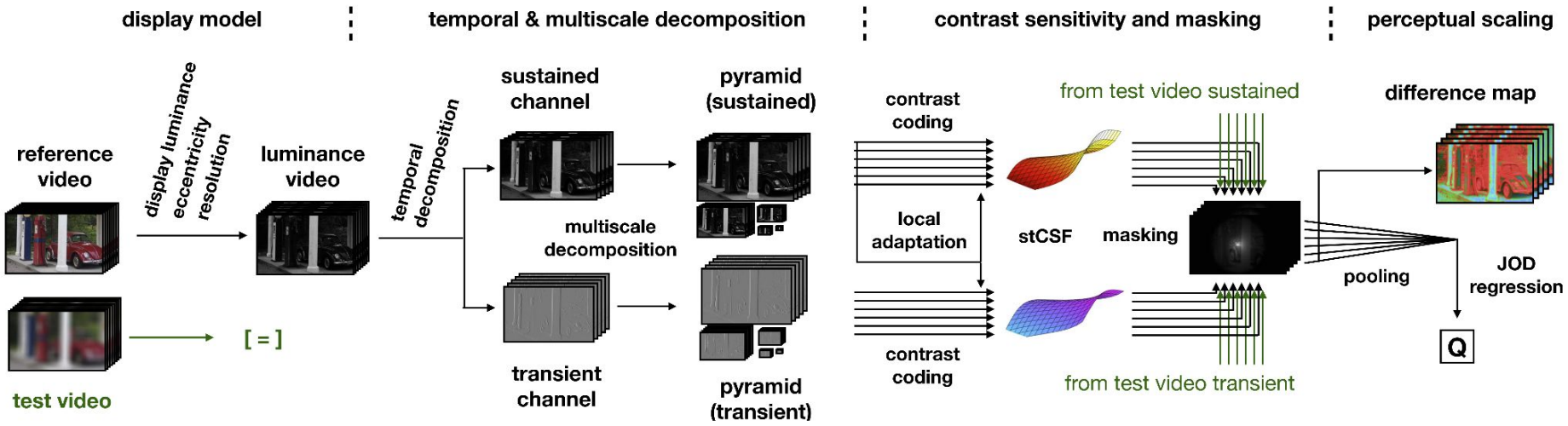
- AVQT = not disclosed yet, but claimed to align well to MOS by internal testing by Apple

(+) supported by Apple with binary library, fast execution

(+) supports beyond 8-bit content, viewing distance adaptation, up to 4K resolution

(–) not many studies so far, no open-source implementation

Quality Metrics: Temporal Aspects – FovVideoVDP



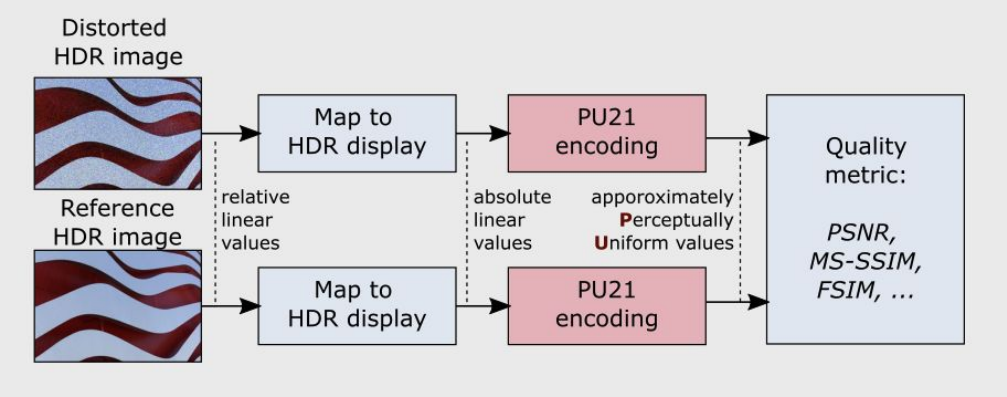
[Mantiuk et al., ACM Trans. Graphics, 2021]

(+) accounts for peripheral acuity, models change over time+visual field

(+) works with SDR and HDR content, Matlab and PyTorch code available

(–) not widely tested so far, may be complex to run at scale

Quality Metrics: HDR-focused PU-21



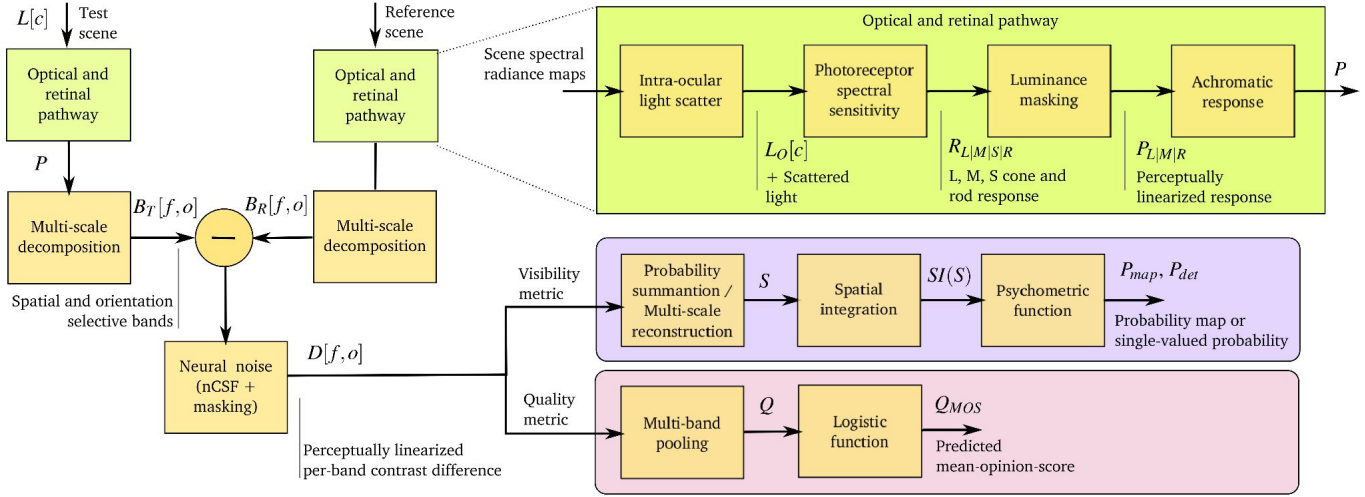
[Mantiuk et al., https://github.com/gfxdisp/pu21]

(+) Generic technique for HDR mapping, any metric can be used subsequently

(+) Code is available

(–) May not apply for all use cases

Quality Metrics: HDR-focused HDR-VDP 2/3



[Mantiuk et al., VDP2, ACM Trans. Graphics, 2011, https://sourceforge.net/projects/hdrvdp/files/hdrvdp/, new paper in preparation for VDP3]

(+) Incorporates temporal, scaling, and several masking properties

(+) Code is available, works for HDR content

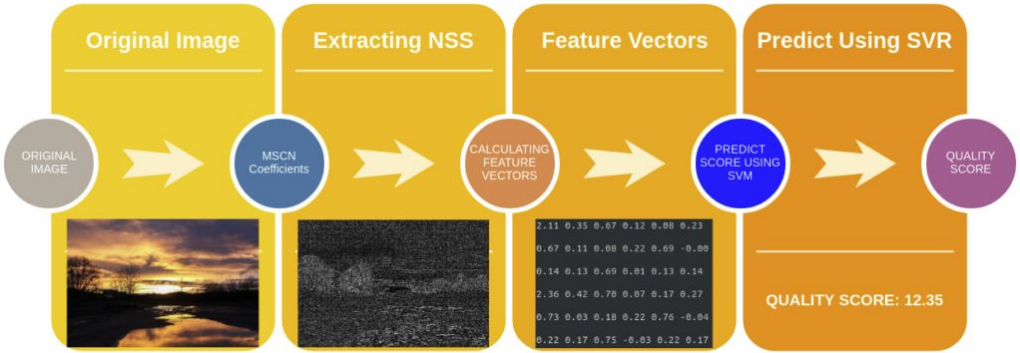
(–) May not apply for all use cases

Quality Metrics: Non-Reference based

- NIQE, BRISQUE
- p.1204
- Other
- AI-based

NR Quality Metrics: BRISQUE, NIQE

- Collect undistorted natural images
- Divisive normalization → norm. image $\mathbf{I}$ → feature vectors (FVs) → Gaussian fits to FVs
- BRISQUE: Use svr to fit feature vectors to MOS corresponding to certain distortion type(s)
- NIQE: (i) Fit multi-variate Gaussian (MVG) model to BRISQUE features
  (ii) Measure deviation from the MVG fit on select local patches



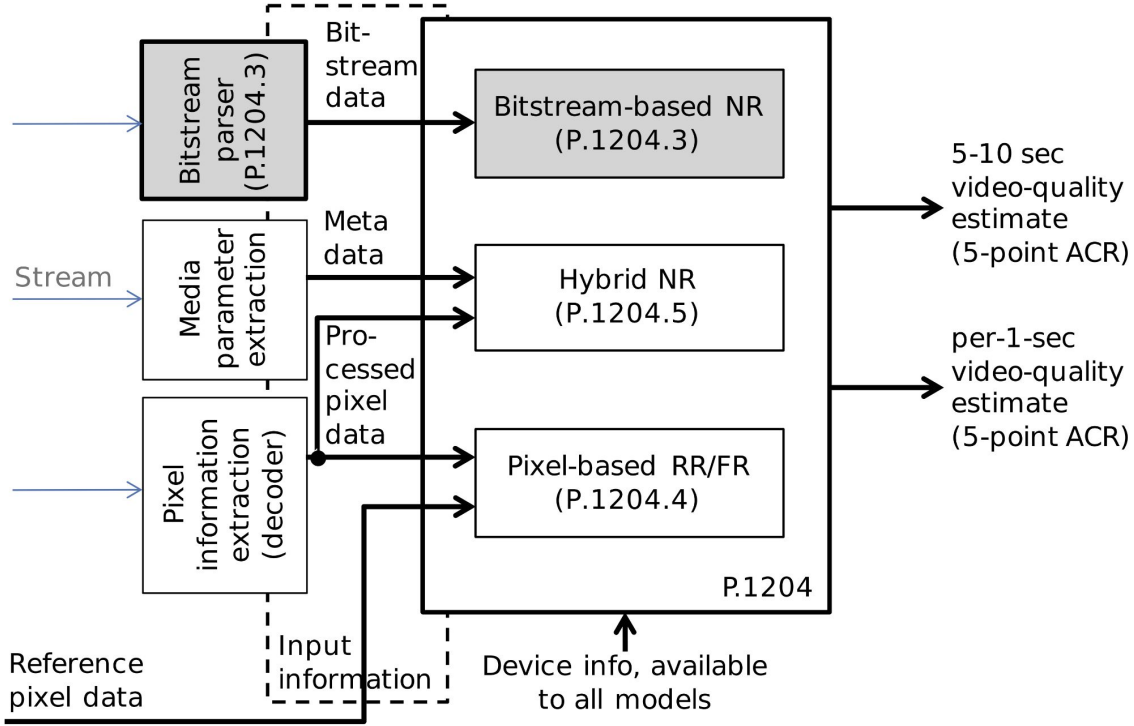
[https://learnopencv.com/image-quality-assessment-brisque/]

$$\mathbf{H} = \mathbf{I}_{0,0} \odot \mathbf{I}_{0,+1}$$
$$\mathbf{V} = \mathbf{I}_{0,0} \odot \mathbf{I}_{+1,0}$$
$$\mathbf{D}_1 = \mathbf{I}_{0,0} \odot \mathbf{I}_{+1,+1}$$
$$\mathbf{D}_2 = \mathbf{I}_{0,0} \odot \mathbf{I}_{+1,-1}$$

(+) Widely available, including within Matlab, fast and easy to use

(–) Will not be accurate enough for many real-world applications

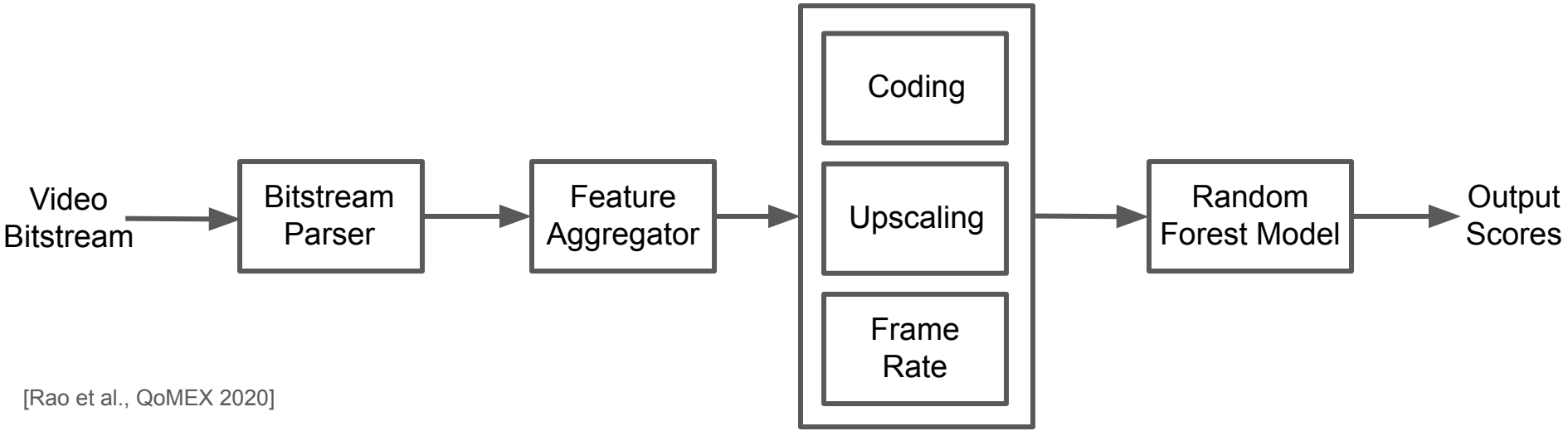
Bitstream-based Quality Metrics: p.1204



[Raake et al. IEEE Access 2020]

NR Quality Metrics: p.1204.3

- Designed to work with compressed-domain content



[Rao et al., QoMEX 2020]

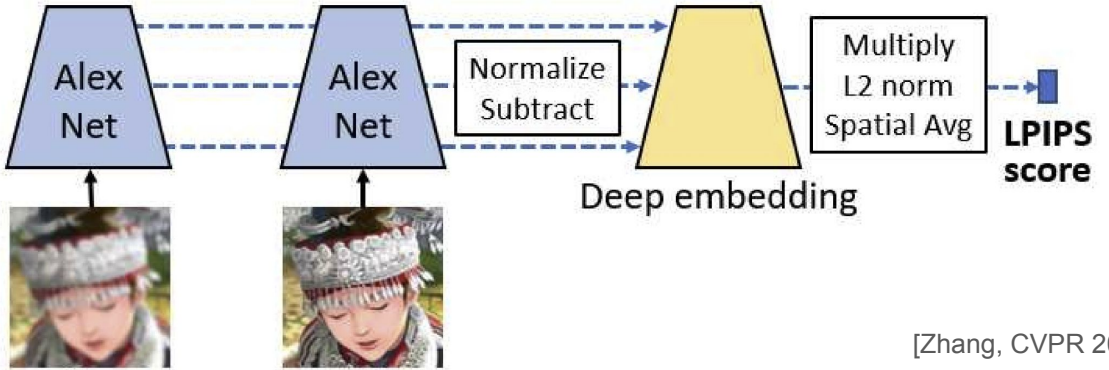
(+) Fast, encoding-standard specific, accurate for some use cases, open-source tools available
(+) Unlike VMAF and other metrics, it is an NR assessment process, scores scale well
(–) Will not support some encoding formats
(–) It may not be as accurate as reference-based quality assessment

NR Quality Metrics: Other

- BLIINDS, DIIVINE
- NORM
- PSTR-PXNR
- AI-based

AI-based FR Quality Metrics: LPIPS

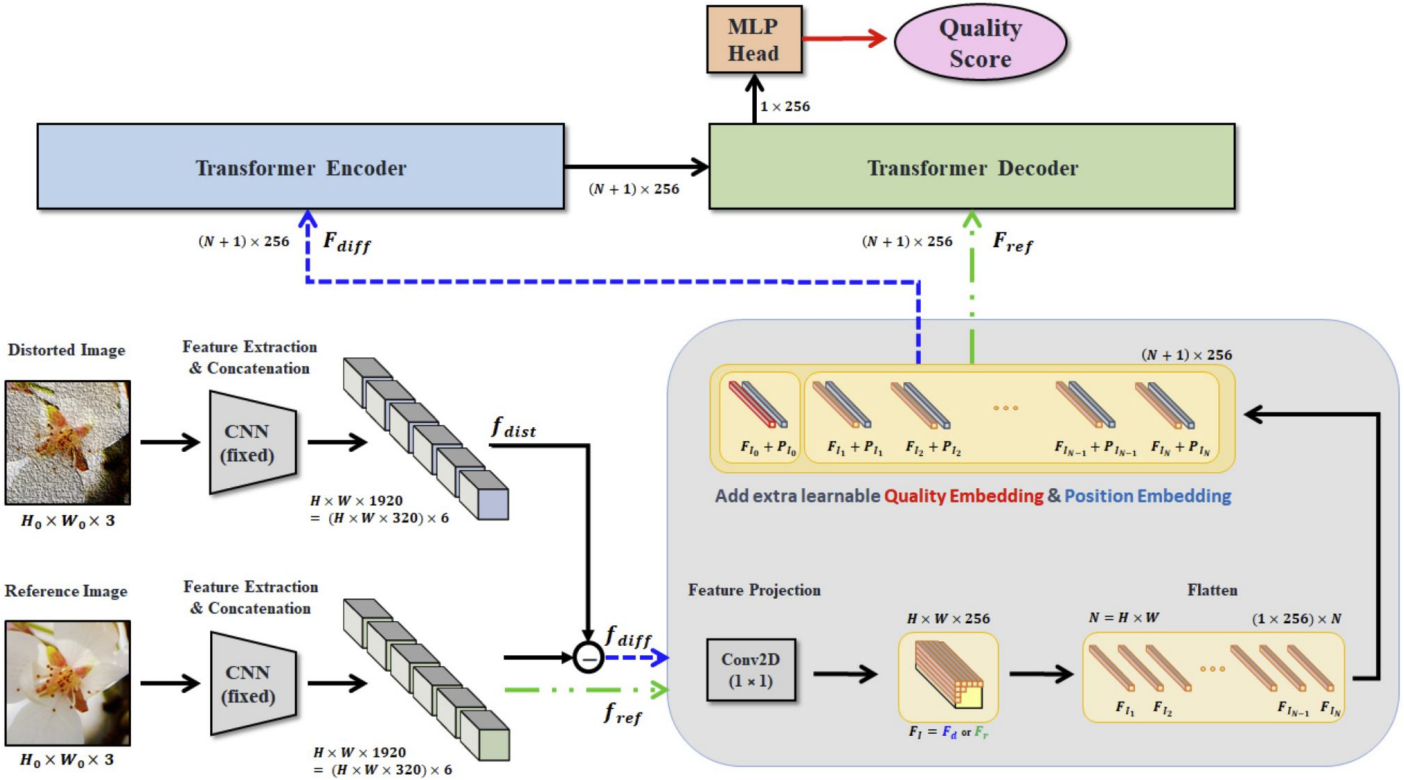
- More perception-oriented, less distortion-oriented



[Zhang, CVPR 2018, Jo, CVPRW 2020]

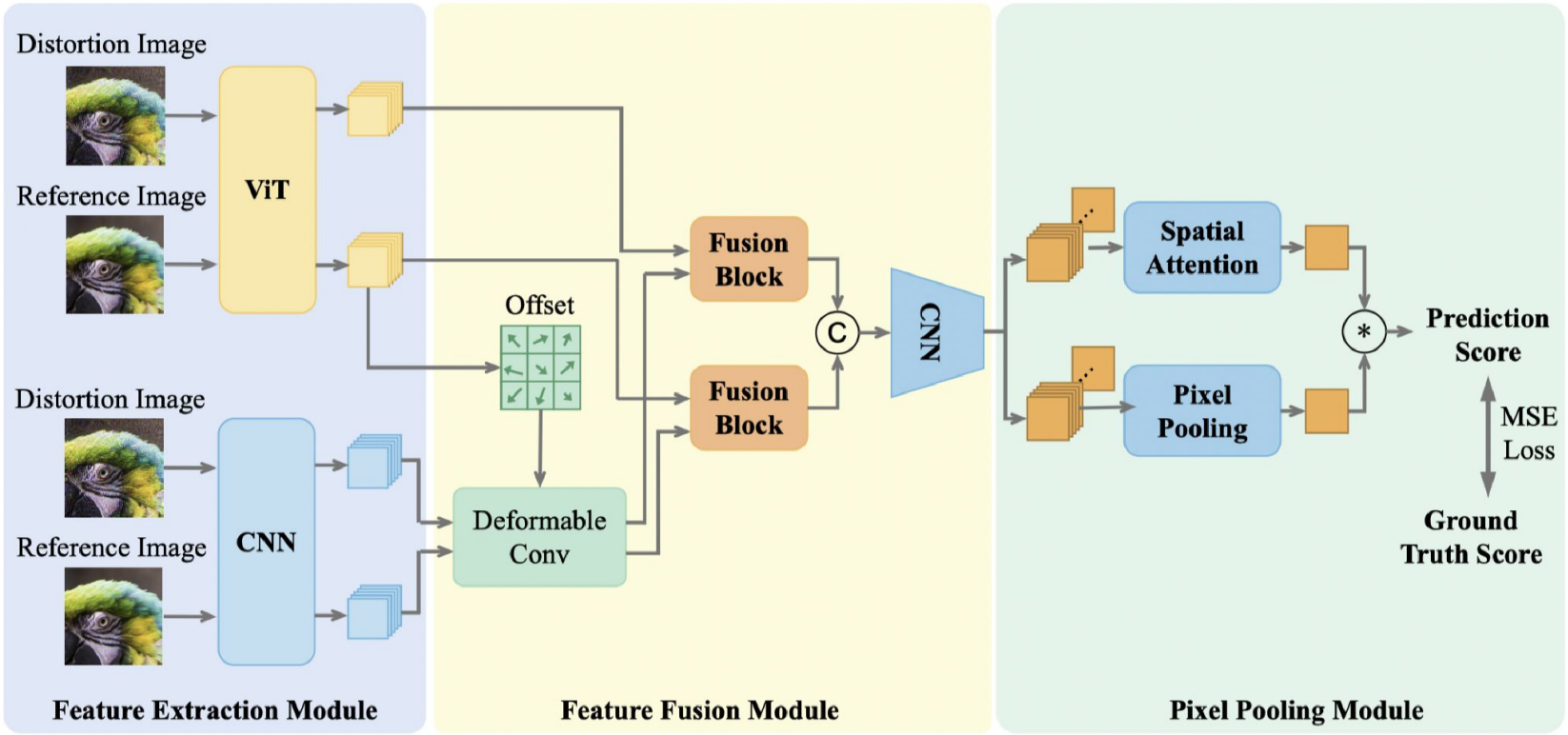
(+) More invariant to imperceptible translation or geometric distortion
(+) Quality range stretches well, well supported by libraries (e.g., PyTorch)
(–) Can fail to detect some distortions
(–) Can be slow to run

AI-based FR Quality Metrics: Transformer-based IQT



[Cheon, Image quality assessment with transformers, CVPRW 2021]

AI-based NR Quality Metrics:



[Perceptual IQA Challenge, NTIRE CVPRW 2022]

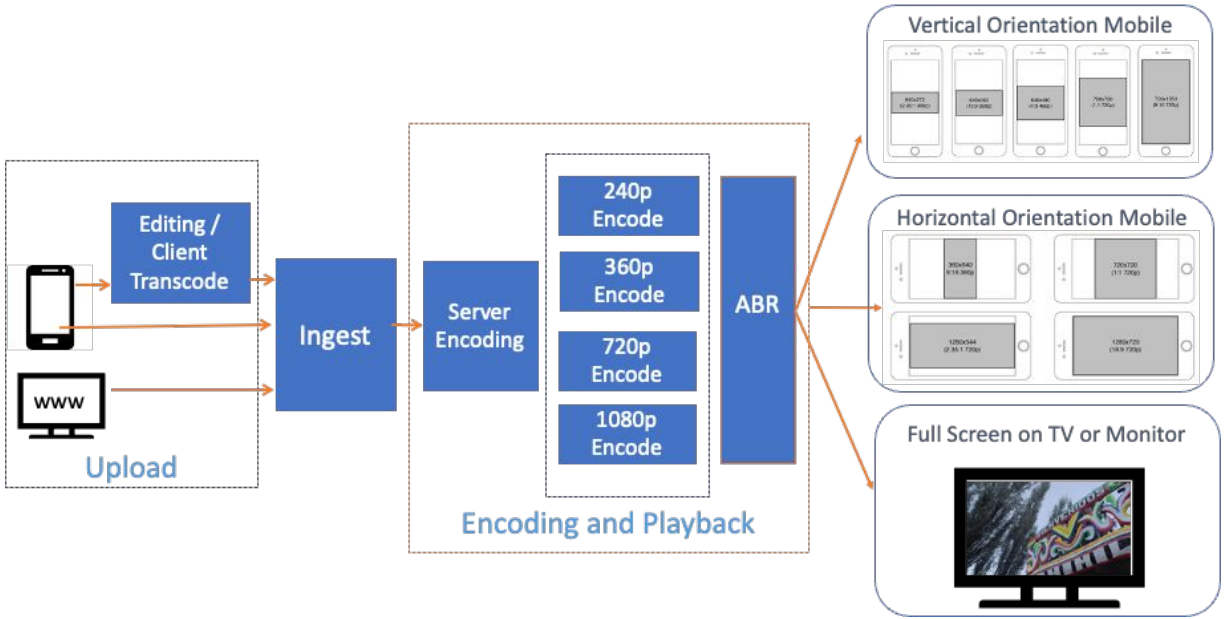
Quality Metrics: NTIRE-2022 Perceptual IQA Challenge

Rank	Team Name	Author/Method	PIPAL-NTIRE22-Test		
			Main Score	SRCC	PLCC
Track 1: Full-Reference IQA					
1	THU1919Group	shanshan	1.6511	0.8227	0.8284
2	Netease OPDAI	CongHeng.	1.6422	0.8152	0.8271
3	KS	JustTryTry	1.6404	0.8170	0.8235
4	JMU-CVLab	burchim	1.5406	0.7659	0.7747
5	Yahaha!	FLT	1.5375	0.7654	0.7722
6	debut_kele	debut	1.5006	0.7372	0.7634
7	Pico Zen	Komal	1.4504	0.7129	0.7375
8	Team Horizon	tensorcat	1.4032	0.7006	0.7027
	Baselines	IQT (NTIRE-21 Winner)	1.5884	0.7895	0.7989
		LPIPS-Alex	1.1369	0.5658	0.5711
		LPIPS-VGG	1.2278	0.5947	0.6331
		DISTS	1.3422	0.6548	0.6873
		SSIM	0.7530	0.3615	0.3915
		PSNR	0.5263	0.2493	0.2769
Track 2: No-Reference IQA					
1	THU_IIGROUP	THU_IIGROUP	1.4436	0.7040	0.7396
2	DTIQA	EvaLab.	1.4367	0.6996	0.7371
3	JMU-CVLab	nanashi	1.4219	0.6965	0.7254
4	KS	JustTryTry	1.4066	0.6808	0.7257
5	NetEase OPDAI	wanghao1003	1.3902	0.6705	0.7196
6	Withdrawn submission	anonymous	1.1828	0.5760	0.6068
7	NTU607QCO-IQA	mrchang87	1.1117	0.5269	0.5848
	Baselines	NIQE	0.1418	0.0300	0.1118
		MA	0.3978	0.1737	0.2242
		PI	0.2764	0.1234	0.1529
		Brisque	0.5722	0.2695	0.3027

Tutorial Outline

- Video streaming, distortion, perception, quality assessment
- Quality metrics and subjective quality assessment
- Example use cases at scale
- Tools
- Future of quality assessment

Use Cases At Scale: End To End Quality Monitoring



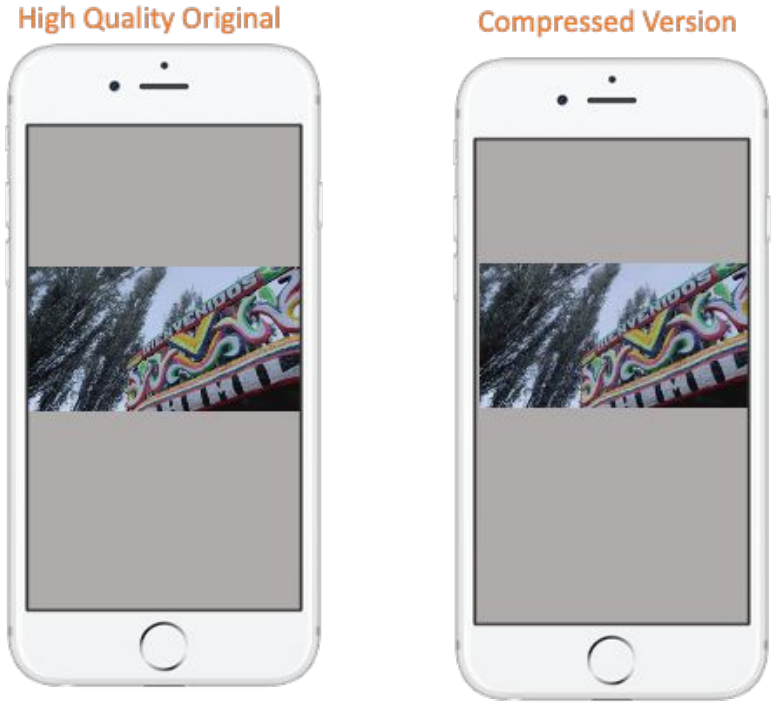
Shankar L Regunathan, et al. 2020. Efficient measurement of quality at scale in Facebook video ecosystem. In Applications of Digital Image Processing XLIII, Vol. 11510. SPIE, 69–80.

Use Cases At Scale: End To End Quality Monitoring

High Quality Original

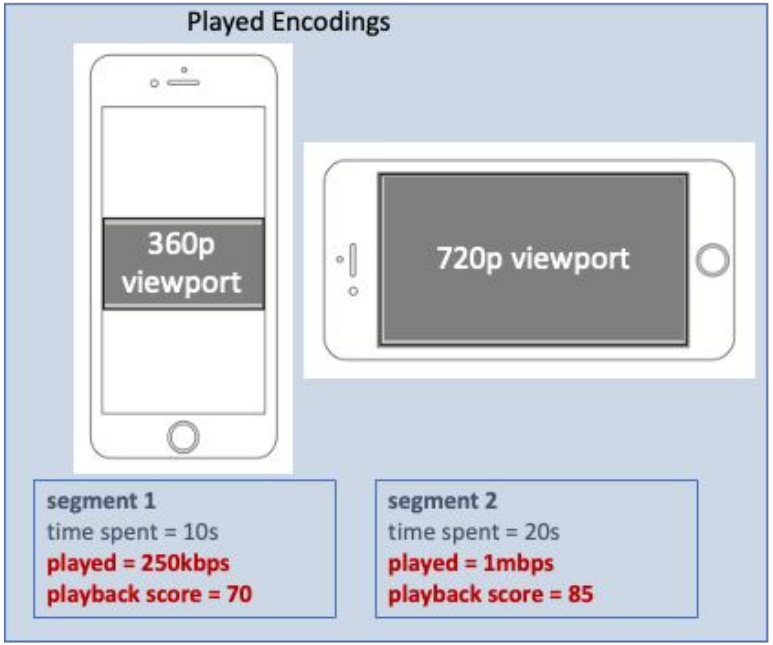
Compressed Version

Use Cases At Scale: End To End Quality Monitoring



Artifacts are harder to perceive in smaller viewport

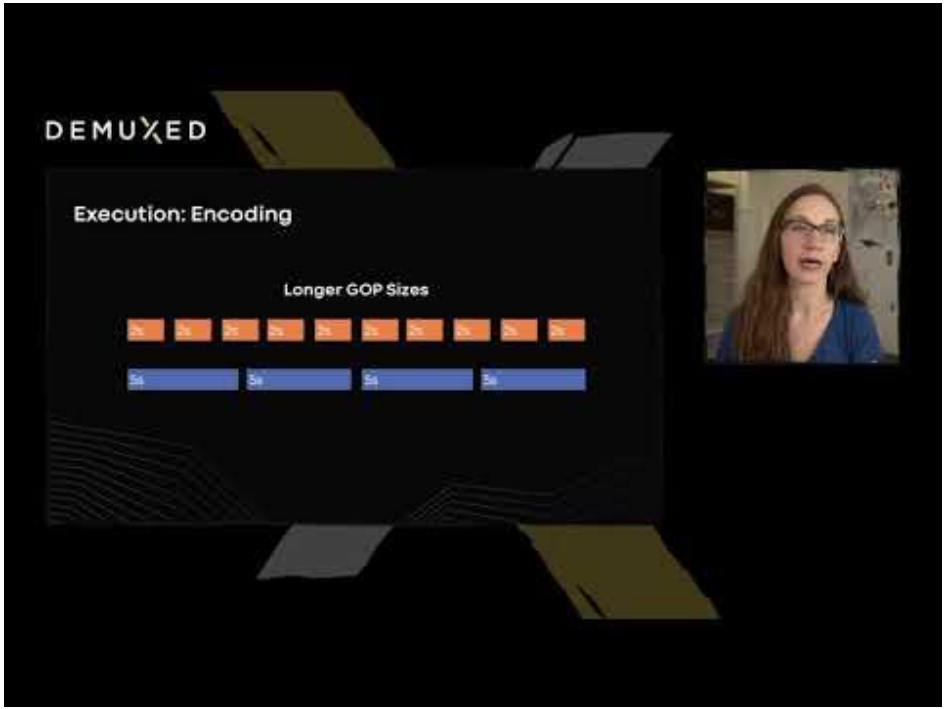
Use Cases At Scale: End To End Quality Monitoring



Encoding	MOS@360p Viewport	MOS@720p Viewport
250kbps	70	40
1mbps	95	85

$$\text{Weighted Quality Score} = \frac{10 \times 70 + 20 \times 85}{10+20} = 80$$

Use Cases At Scale: Optimize Video Experience



Denise Noyes - Providing better video experiences for the next billion users, Demuxed 2020

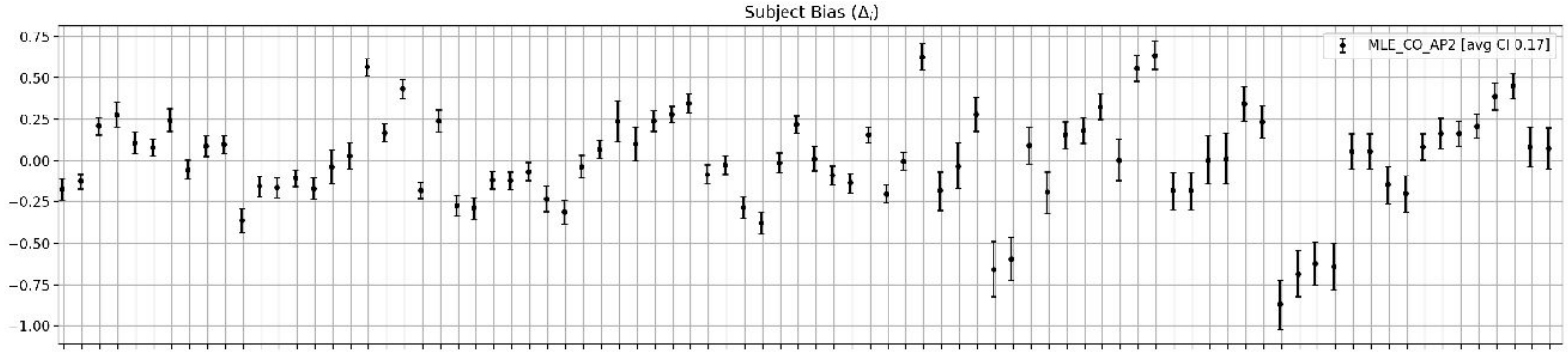
https://www.youtube.com/watch?v=hKHtGTRdtjI

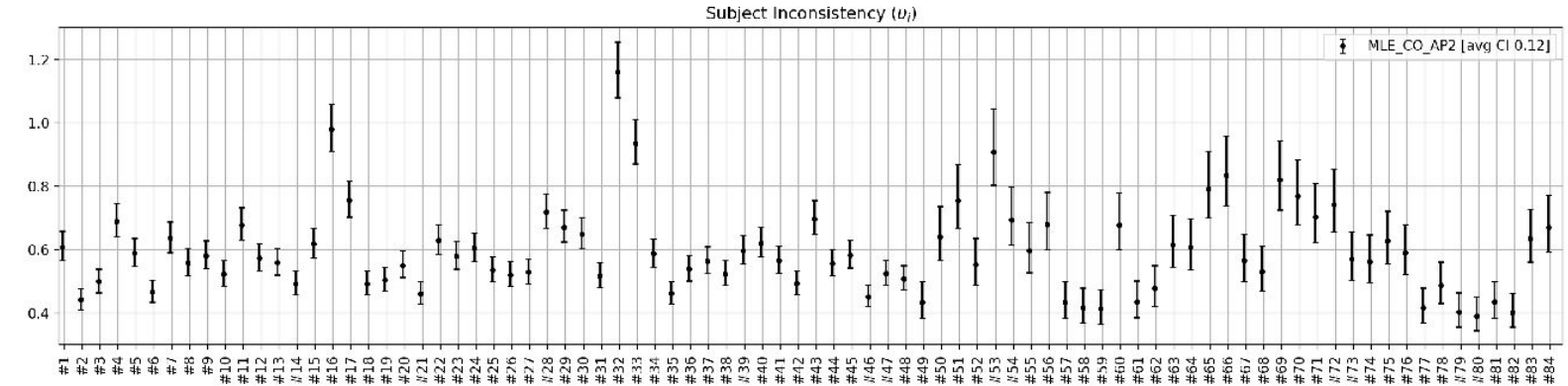
Use Cases: Next-gen Video Compression/Rendering

Use Cases: P.910 ACR-HR of Advanced Encoding Tools

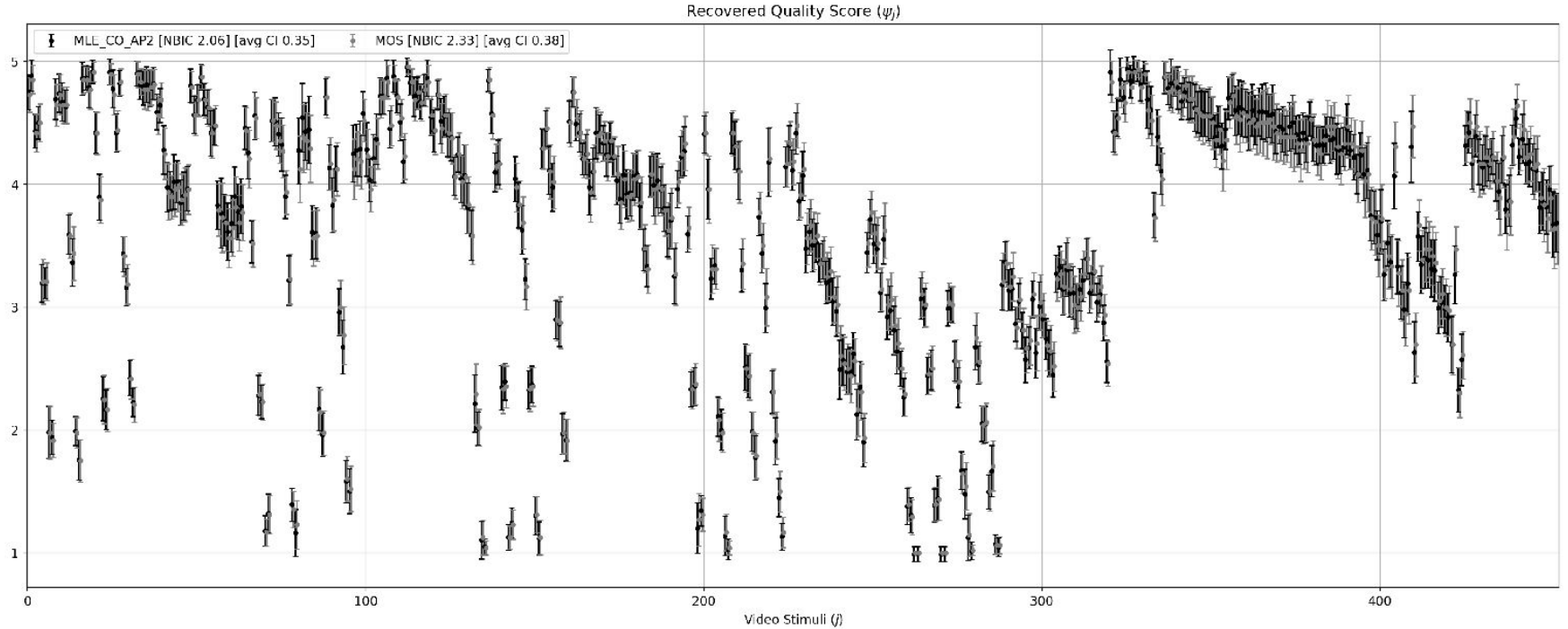
Setup Component	What was Used	Further Details on Settings	Comments
Encoders	AVC x264 (Lavc58.134.100 libx264) WEBM VP9 (v1.10.0-48-g4ec84326c)	• 1080p, 720p, 540p, 360p, 216p (only underlined done for post-processing) • Per resolution: AVC preset=veryslow, CRF={22,30,38,46} (medium used for post-processing) • Per resolution: VP9 preset=0, CRF={32,38,44,46,48,50,52,54,56,58,60} (underlined CRFs done for 720p & 540p, preset=5 used for post-processing)	• The slowest preset of each encoder was used for preprocessing, faster presets for post-processing • Constant-CRF encoding ensures quality remains consistent, no effects from rate control algorithms • The range of CRFs ensures the full quality range of relevance to each resolution & application is sampled • All lower resolutions were upscaled to 1080p for viewing using FFmpeg Lanczos-5
Content and test conditions	AV2 CTC content https://media.xiph.org/video/aomctc/test_set/ P.910 ACR-HR standard test conditions applied	• 3H distance, controlled lighting, same screen conditions for all tests • Ratings from 1-5 • Raters were briefed on task and how to use the quality scaling	• All content replayed at 25fps, 1080p@50Hz TV screen, all TV filters were off • 21 sequences at 1080p resolution (8bit) used, comprising a mixture of entertainment, sports, UGC, gaming, web browsing, and artistic content (16 sequences for post-processing)
Raters and data processing	• 48 raters for preprocessing (the underlined VP9 CRFs had 36 additional raters) • 24 raters for post-processing • The SUREAL package was used for post-processing	• All raters were screened for color blindness and good eyesight • All 16368 ratings were used	• SUREAL: https://github.com/Netflix/sureal • The full maximum likelihood estimation (MLE) model of SUREAL was used • An MLE fit per codec was carried out and the recovered quality scores were used

P.910 Subject Bias and Inconsistency



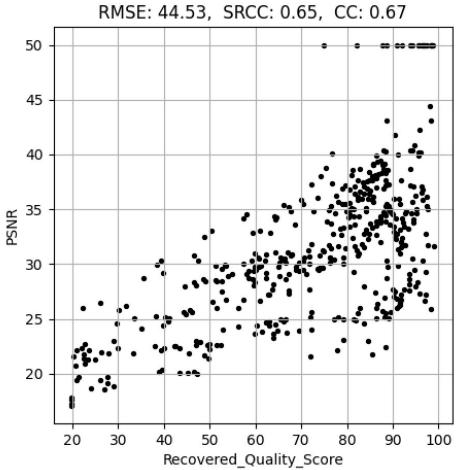


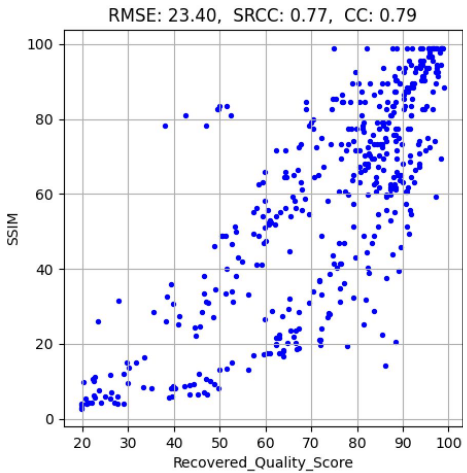
P.910 Recovered Quality Scores

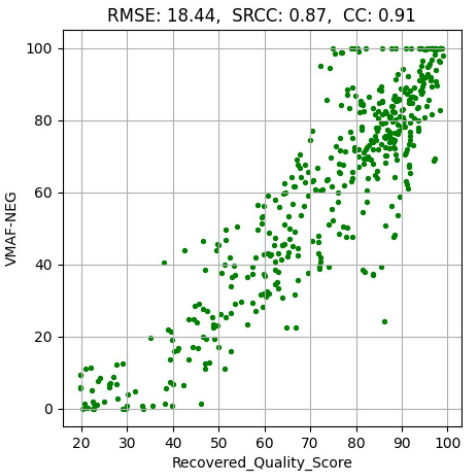


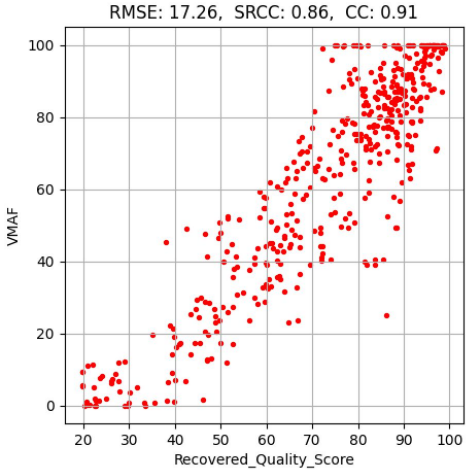
- The Recovered Quality Scores (RQS) span the entire quality range and are adjusted according to bias, uncertainty and inconsistency based on SUREAL's methodology

P.910 Metrics vs. RQS



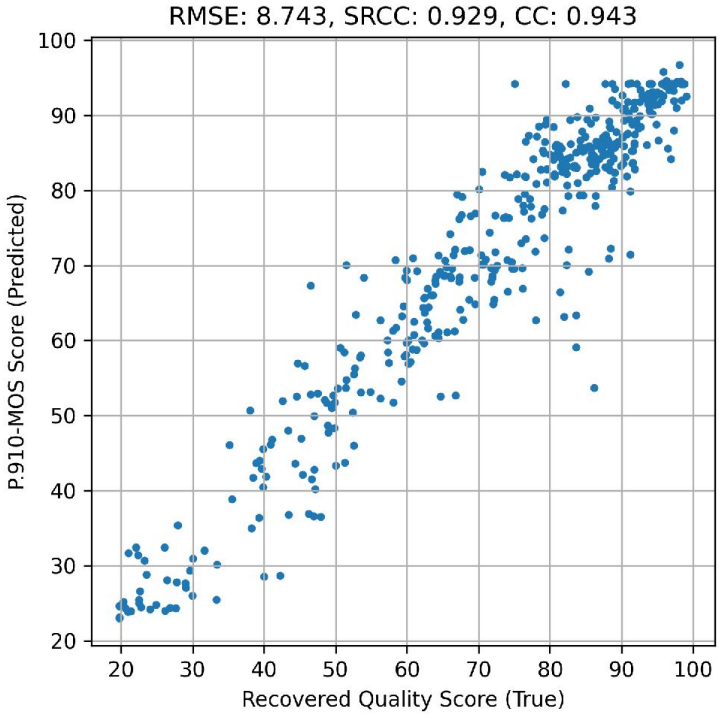






- VMAF-NEG and VMAF are well aligned to Recovered Quality Scores, with correlation of 91%

P.910 SVR Results



- Scatter plot of SVR with v=0.5 (proportion of support vectors vs. total samples), γ=0.85 (radius of RDF), C=1 (regularization term) predicted scores vs recovered quality scores

Use Case: Reduce VMAF complexity for use at scale

- VMAF has state of the art model performance
- However it is expensive to compute at scale
- Can we create an alternative model with less complex features?

Use Case: Reduce VMAF complexity for use at scale

- Both VIF and DLM are multi-scale methods
- But they do not reuse the same pyramid
- VIF pyramid is expensive (17x17, 9x9, 5x5, 3x3)
- DWT is cheaper, 4x4 for db2

Use Case: Reduce VMAF complexity for use at scale

- Unifying all features on the same wavelet transform can reduce complexity by ~4x
- Result is FUNQUE: Fusion of Unified Quality Evaluators
- To be presented ICIP 2022

Tutorial Outline

- Video streaming, distortion, perception, quality assessment
- Quality metrics and subjective quality assessment
- Example use cases at scale
- Tools
- Future of quality assessment

Tools: SSIM

Papers on mathematical properties:

- Z. Wang, et al. “Multiscale structural similarity for image quality assessment,” Proc. IEEE Asilomar Conf. on Signals, Systems & Computers, 2003.
- A. Hore, and D. Ziou, “Image quality metrics: PSNR vs. SSIM,” Proc. IEEE 20th Int. Conf. Pattern Recognition (pp. 2366-2369), Aug. 2010.
- D. Brunet, et al. “On the mathematical properties of the structural similarity index,” IEEE Trans. on Image Processing, vol. 21, no. 4, pp. 1488-1499, 2011.
- Y. Reznik, “Another look at SSIM image quality metric," Proc. Picture Coding Symposium (PCS 2022), San Jose, CA, 7-9 December 2022.

Tools: SSIM

Papers on fits to MOS:

- A. K. Venkataramanan, “A Hitchhiker’s guide to structural similarity,” IEEE Access, 9 (2021): 28872-28896.
- S. L Regunathan, et al., “Efficient measurement of quality at scale in Facebook video ecosystem,” Proc. SPIE Applications of Digital Image Processing XLIII, Vol. 11510, 2021.

SSIM resources:

- libvmaf, FFmpeg, Matlab, Scikit-Video in Python, PyTorch, Tensorflow,...

Tools: VMAF Enhancements

Papers on VMAF extensions:

- VMAF-NEG for enhancement gain limit
- CAMBI for banding artifacts in video
- M. Utke, et al. “NDNetGaming-development of a no-reference deep CNN for gaming video quality prediction,” Multimedia Tools and Applications (2020).
- M. Orduna, “Video multimethod assessment fusion (VMAF) on 360VR contents,” IEEE Trans. Consumer Electronics, vol. 66, no. 1, pp. 22-31, 2019.
- D.Ramsook, et al. “A differentiable estimator of VMAF for video,” Proc. Picture Coding Symposium (PCS). IEEE, 2021.

VMAF resource:

- libvmaf, https://github.com/Netflix/vmaf

Tools: NR metrics

- P.1204.3

https://github.com/Telecommunication-Telemedia-Assessment/bitstream_mode3_p1204_3

- NTIRE 2022 NR competition and VQEG NORM (see next slide)

Tools: VQEG, NTIRE Datasets and Methods

VQEG resources of relevance:

- No-reference metric resources & datasets https://vqeg.org/projects/norm-resources.aspx
- Audiovisual HD quality https://vqeg.org/projects/audiovisual-hd/
- Video datasets https://vqeg.org/video-datasets-and-organizations/
- Publications and software https://vqeg.org/publications-and-software/
- Presentations at meetings https://vqeg.org/meetings-home/

NTIRE competitions:

- Challenges of the 2022 CVPR workshop: https://data.vision.ee.ethz.ch/cvl/ntire22/
  → Perceptual Image Quality Assessment (FR and NR tracks)
  → Super-resolution (efficiency and learning the SR space)
  → Video/multi-frame challenges

Tools: Subjective Quality Assessment

P.910 crowdsourcing and post-processing

- Microsoft repo on crowdsourcing P.910 https://github.com/microsoft/P.910
- Netflix SUREAL https://github.com/Netflix/sureal
- UTexas video quality challenge dataset: https://live.ece.utexas.edu/research/LIVEVQC/index.html
- VQEG SAM group https://vqeg.org/projects/statistical-analysis-methods-sam.aspx

Tutorial Outline

- Video streaming, distortion, perception, quality assessment
- Quality metrics and subjective quality assessment
- Example use cases at scale
- Tools
- Future of quality assessment

Future of Quality Assessment of Video

- Quality assessment of 360-deg video https://www.itu.int/rec/T-REC-P.919-202010-I
- QA of HDR tonemapping https://hal.archives-ouvertes.fr/hal-02612844/document
- QA for frame-rate conversion
- Metrics for 3D or VR/rendered content
- Crowdsourced quality assessment