TOWARDS JOINT FRAME-LEVEL AND MOS QUALITY PREDICTIONS WITH LOW-COMPLEXITY OBJECTIVE MODELS

icassp 2022 Singapore #1696 icassp 2022 Singapore

Tencent Media Lab **Joel Jung, Alexandre Giraud, Meijia Song, Songnan Li, Xiang Li, Shan Liu** Tencent Media Lab

1. CONTEXT & CONTRIBUTIONS

Keywords / context:

- Cloud Gaming
- Automatic quality evaluation using bitstream derived features (qp, mv, …)
- Very low complexity and very low delay, no access to decoded pixels

Framework: P.BBQCG: Parametric bitstream-based Quality Assessment of Cloud Gaming Services
Work item of ITU-T Study Group 12 Question 14

New family of metrics, maximizing classical performance indicators vs MOS, and jointly maximizing a new performance indicator reflecting ability of metrics to detect sudden abrupt quality changes at frame level, that occur frequently in gaming content

Contributions:

1. Evaluation of existing metrics on heterogenous gaming dataset (different codecs, bitrates, …)
2. Design of three new very low-delay and very low-complexity learning-based models
3. Proposition of a new frame level performance indicator to consider gaming content characteristics
4. Proposition of a new training approach to optimize models on all performance indicators

2. GAMING DATASETS

KUGVD dataset

- Kingston University
- 30 videos, 6 games, 1080p@30fps
- H.264, 600 kbps to 4 Mbps
- Lab tests, ACR

CGVDS dataset

- TU Berlin
- 39 videos from 13 games, 1080p@60fps
- H.264 (NVENC), 2 to 6 Mbps

TGDS dataset

- Tencent Media Lab
- 170 videos from 34 scenes of Fortnite, Blade&Soul, Path of Exile, League of Legend, The Witcher
- 1080p@60fps
- H.264 (proprietary), low-delay, 6 to 30 Mbps
- Crowd-sourcing tests, 64 workers
- 5 grade scale, ACR
- Strict outlier removal process applied
- Confidence interval: 0.33

→ **3 datasets merged:** 239 PVS and 53 different scenes
Difficulty increased / Realism increased

3. STATE OF THE ART EVALUATION

7 state of the art metrics tested (references in the paper):

- PSNR, SSIM, VMAF: complex, full reference, with access to decoded pixels
- NDNetGaming: complex, no reference, NN based, trained on gaming content
- DBCNN: complex, no-reference, DNN based
- P.1203.1, P.1204.3: low-complexity bitstream based models

Dataset split: training (186 PVS) - testing (53 PVS)

Linear mapping applied for RMSE computation (based on training set) [ITU-T P.1401]

	PSNR	SSIM	VMAF	P.1203	P.1204	DBCNN	NDNet
RMSE	0.58	0.56	0.44	0.47	0.46	0.50	0.42
PLCC	0.67	0.68	0.81	0.79	0.80	0.76	0.83
SRCC	0.65	0.78	0.82	0.80	0.80	0.74	0.82

→
- **Low performance of PSNR and SSIM**
- **Good performance of other metrics**

especially P.1203.1 and P.1204.3, close to more complex methods

4. THREE NEW MODELS & RESULTS: learning-based / low-complexity / low-delay

VQMCG: Video Quality Metrics for Cloud Gaming: new learning-based models: no-reference, no-access to decoded pixels

increasing complexity / increasing accuracy

- **VQMCG.a:** weighted linear combination of features, weights learnt on training set with gradient descent
- **VQMCG.b:** Support Vector Regression (SVR), supervised learning algo mapping the features space with MOS by finding a hyperplane on the training set
- **VQMCG.c:** Multi-Layer Perceptron (MLP), NN with fully connected layers, weights initialized with Glorot, trained with back-propagation with Adam optimizer (4 layers with 100, 50, 25 and 10 neurons, activated by a ReLU function)

Features:

VQMCG based on features extracted or computed from the parsed bitstream:

Qp	Avg nb P macroblocks / frame
Frame size	Avg nb of 8x8 blocks
Bitrate	Avg nb of blocks with 8x8 transform size
Spatial complexity	Avg nb of blocks without frequential transform
	Avg nb of skipped blocks

	VQMCG.a	VQMCG.b	VQMCG.c
RMSE	0.40	0.32	0.29
PLCC	0.87	0.91	0.93
SRCC	0.88	0.91	0.92

to compare with table of part 3

→ **VQMCG.a better than VMAF and NDNetGaming**
VQMCG.b and VQMCG.c outperform existing methods on "classical" indicators

→ Why high performance for the learning-based models? Games (CG) made of repetitive visual characteristics (motion pattern, color diversity, backgrounds, …)
Similar repeating scenes, spatial and temporal similarities: **gaming content adapted to learning-based models**

5. DIFFICULTY OF GAMING CONTENT

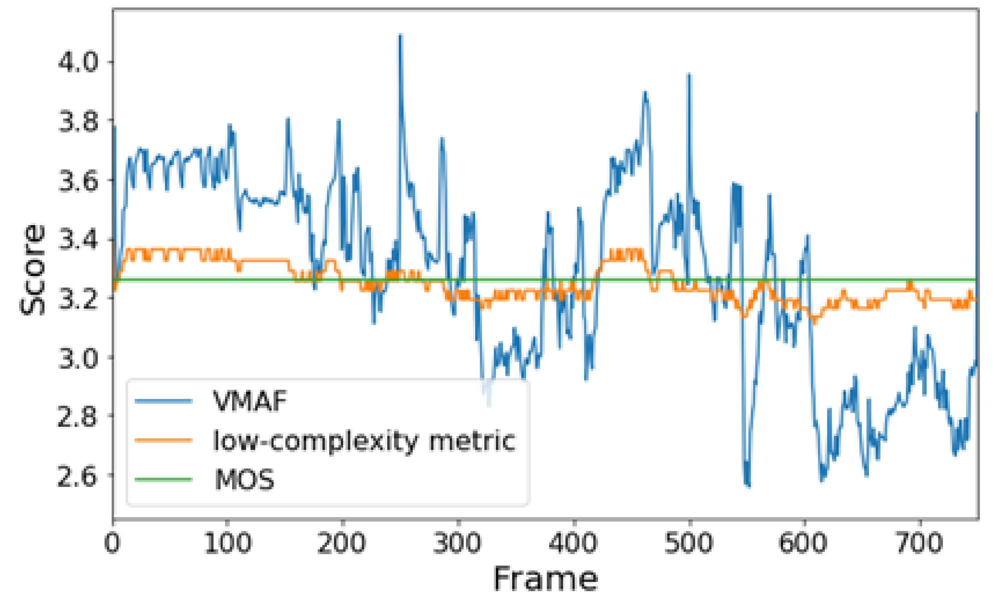
Gaming content characteristics:

- sudden & fast rotations
- explosions

→ high & sudden temporal and spatial intensity changes

↓ codec "stress" (especially low delay): motion prediction issue, intra blocks, …

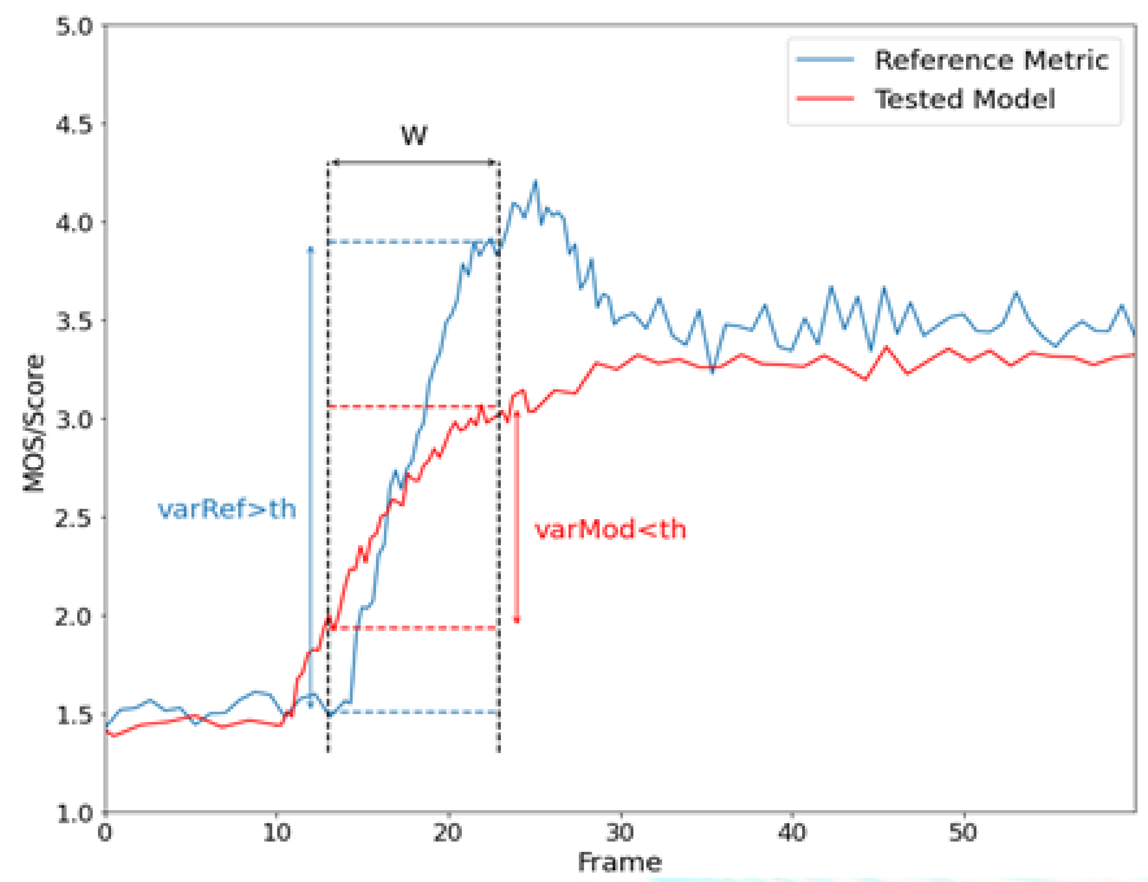
← **large and abrupt/sudden quality changes…**
… not reflected by low-complexity metrics



→ **Achieving good correlation with MOS on segments of several seconds is insufficient**

6. NEW FRAME LEVEL CORRELATION INDICATOR

FVM: Frame Variation Match: measures ability of a model to reflect large and sudden quality variations when a reference metric reports similar kind of variation



FVM: counts % of time when the model has a quality change *varMod* above a threshold *th*, when a reference metric also has a quality change *varRef* above *th*, and in the same direction, in the same window *W*

Reference metric: VMAF

(any reliable full-reference metric can be used)

Matching a full-reference metric with low-complexity models at the frame level is needed

	P.1203.1 / #	P.1204.3 / #	VQMCG.a	VQMCG.b	VQMCG.c
FVM	2% / 36%	4% / 29%	71%	24%	31%

(#) represents frame level modifications of P.1203.1 & P.1204.3

→ **Except for VQMCG.a, FMV indicator is too low!**

7. FINAL RESULTS WITH "GAMING FRIENDLY" LEARNING & CONCLUSION

Proposed training process:

- Reference objective scores used at a frame level
- MOS used as an offset applied to the objective scores

models trained on VMAF shifted, centered on MOS

	VQMCG.a*	VQMCG.b*	VQMCG.c*
RMSE / MOS	0.40	0.26	0.30
PLCC / MOS	0.86	0.94	0.92
SRCC / MOS	0.87	0.93	0.91
FVM / VMAF	36%	51%	50%

→ **FVM vs VMAF @ frame level improved AND other indicators vs MOS preserved**

→ **Excellent correlation on all indicators**