VQEG 2023
San Mateo, CA, USA

Green Blind Visual Quality Assessment for Real-Time Communications

C.-C. Jay Kuo
William M Hogue Professor
University of Southern California

Visual Quality Assessment (VQA)

- Two Application Scenarios
  - Professional video content streaming: both raw and compressed videos are available
  - UGC video streaming and conversational video: no-reference videos are available

- User Generated Content (UGC)
- Capture/display by smartphones

- Multi-party conversational Video
- Low latency requirement

VQA of Professional Video

- Solution: VMAF (Video Multi-Method Assessment Fusion)
- Collaboration between USC and Netflix (2014-2015)
- Received Technology and Engineering Emmy Award (2020)

NETFLIX

VMAF

VQA of UGC & Conversational Video

- Main Challenges
  - No reference
  - Limited computational resources (i.e., memory & power consumption)
  - Low latency for interaction

- Solution:
  - Lightweight machine learning solution

About Me

- C.-C. Jay Kuo
- William M. Hogue Professor and Distinguished Professor at USC
- Director of Media Communication Lab (MCL)
- Fellow of AAAS, ACM, IEEE, NAI and SPIE.
- Academician, Academia Sinica (Taiwan)
- Publications: 15 books, 30 patents, 340 journal papers, 1000 conference papers

Industrial Collaboration (with 70+ Companies)

Collaboration with Meta (2022)

January 2022 - March 2022	March 2022 - October 2022	October 2022 - December 2022
Blind image quality assessment (BIQA)	**Blind video quality assessment (BVQA)**	**Reports and demos**
• Design quality-aware feature extractor for BIQA. • Adopt regressors for perceptual quality score predictions. • Conduct experiments on benchmark BIQA datasets. • Time & memory analysis.	• Extract features for I-frames. • Include temporal information into our system, including motions, residuals, and etc. • Conduct experiments on benchmark BVQA datasets. • Time & memory analysis. • Refining BIQA.	• Wrap up the results and produce final reports and demos. • Code organization and documentation.

- Introduction
- **Green Blind Image Quality Assessment**
- Green Blind Video Quality Assessment
- Takeaway

Challenges

- Datasets
  - Subjective scores are expensive to obtain
  - Authentic datasets contain mixed and complex distortions
- Existing methods
  - Conventional methods
    - Hand-crafted features
    - Lack of expressiveness for user-generated images/videos
  - Deep learning methods
    - Huge models pre-trained on large datasets
    - High latency and computing complexity for mobile or edge devices

Our approach

- Green learning method [1]
  - Lightweight model without backward propagation
  - Low latency
  - Low computational resources
  - Reasonable performance

[1] C.-C. J. Kuo and A. M. Madni, “Green learning: Introduction, examples and outlook,” Journal of Visual Communication and Image Representation, vol. 90, p. 103685, 2023

GreenBIQA – Exemplary Images

- Mean opinion scores (MOS)
- 1: Bad, 2: Poor, 3: Fair, 4: Good, 5: Excellent

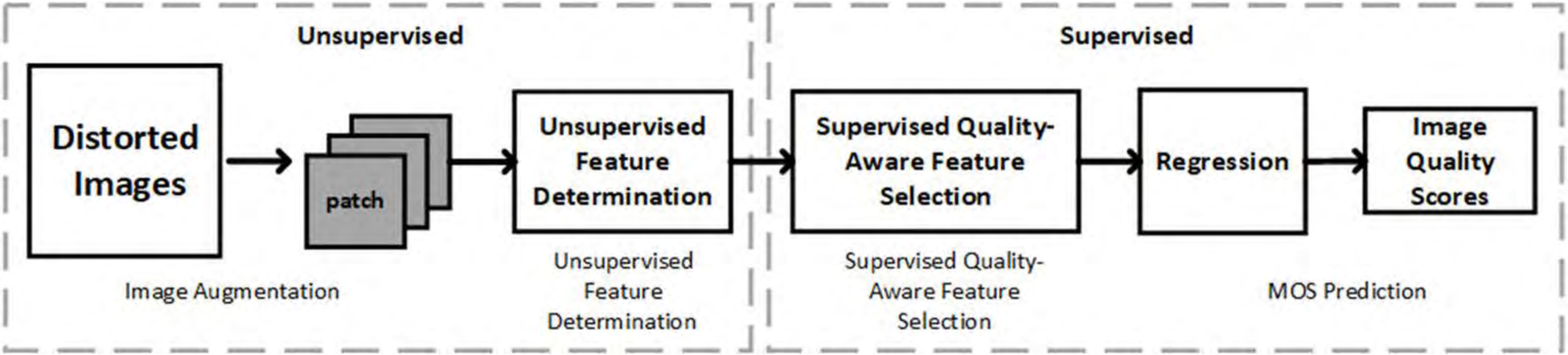
MOS: 3.63

MOS: 3.72

MOS: 4.21

MOS: 1.78

GreenBIQA - Pipeline



- Increase the number of training samples
- Capture more image patterns

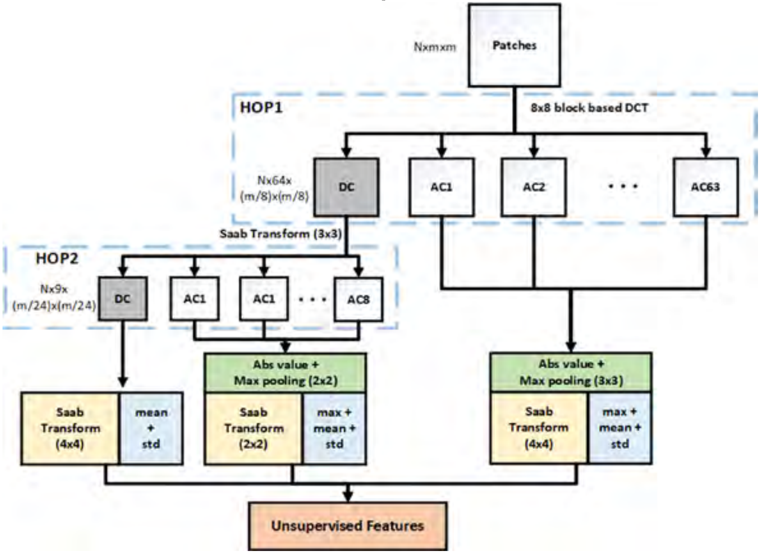
- Generic image features

- Quality-aware image features

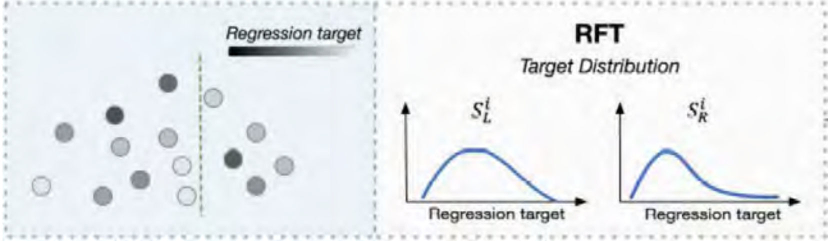
- Quality score prediction
- Decision ensemble

GreenBIQA - Image features

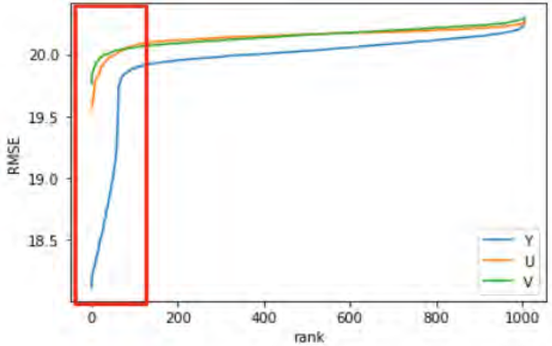
- Multi-hop feature determination
  - Two-hop is sufficient for quality assessment → parameter efficient



- Quality-aware feature selection



Discriminant features for BIQA



GreenBIQA – Prediction Results

MOS: 3.63
Prediction: 3.65
Accurate

MOS: 3.72
Prediction: 3.81
Accurate

MOS: 4.21
Prediction: 3.75
Under-estimate

MOS: 1.78
Prediction: 2.68
Over-estimate

GreenBIQA - Performance Benchmarking

BIQA Method	Dataset: LIVE-Challenge		Dataset: KonIQ-10K		Model Size(MB)	GFLOPs	KFLOPs/pixel
	SROCC	PLCC	SROCC	PLCC			
NIQE	0.455	0.483	0.531	0.538	-	-	-
BRISQUE	0.608	0.629	0.665	0.681	-	-	-
CORNIA	0.632	0.661	0.780	0.795	7.4 (4.07×)	-	-
HOSA	0.661	0.675	0.805	0.813	**0.23 (0.13×)**	-	-
BIECON	0.595	0.613	0.618	0.651	35.2 (19.34×)	0.088 (2.6×)	85.94 (126.8×)
WaDIQaM	0.671	0.680	0.797	0.805	25.2 (13.85×)	0.137 (4.0×)	133.82 (197.4×)
NIMA(Inception-v2)	0.637	0.698	-	-	37.4 (20.55×)	4.37 (128.5×)	87.10 (128.5×)
PQR	0.857	**0.882**	0.880	0.884	235.9 (129.62×)	-	-
DBCNN	0.851	0.869	0.875	0.884	54.6 (30.00×)	16.5 (485.3×)	328.84 (485.1×)
HyperIQA	**0.859**	**0.882**	**0.906**	**0.917**	104.7 (57.53×)	12.8 (376.5×)	255.10 (376.3×)
GreenBIQA (Ours)	0.801	0.809	0.858	0.870	1.82 (1×)	**0.034 (1×)**	**0.678 (1×)**

Model Performance (Correlations with MOS)

Model Memory Efficiency

Computational Complexity

Introduction

Green Blind Video Quality Assessment

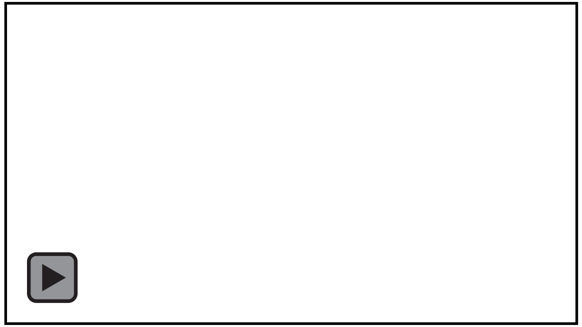
Green Blind Image Quality Assessment

Takeaway

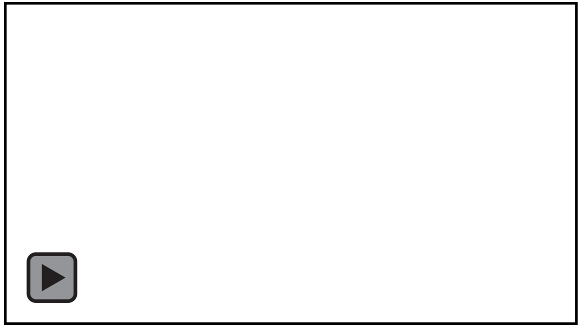
GreenBVQA – Exemplary Videos

MOS: 2.61

MOS: 3.60



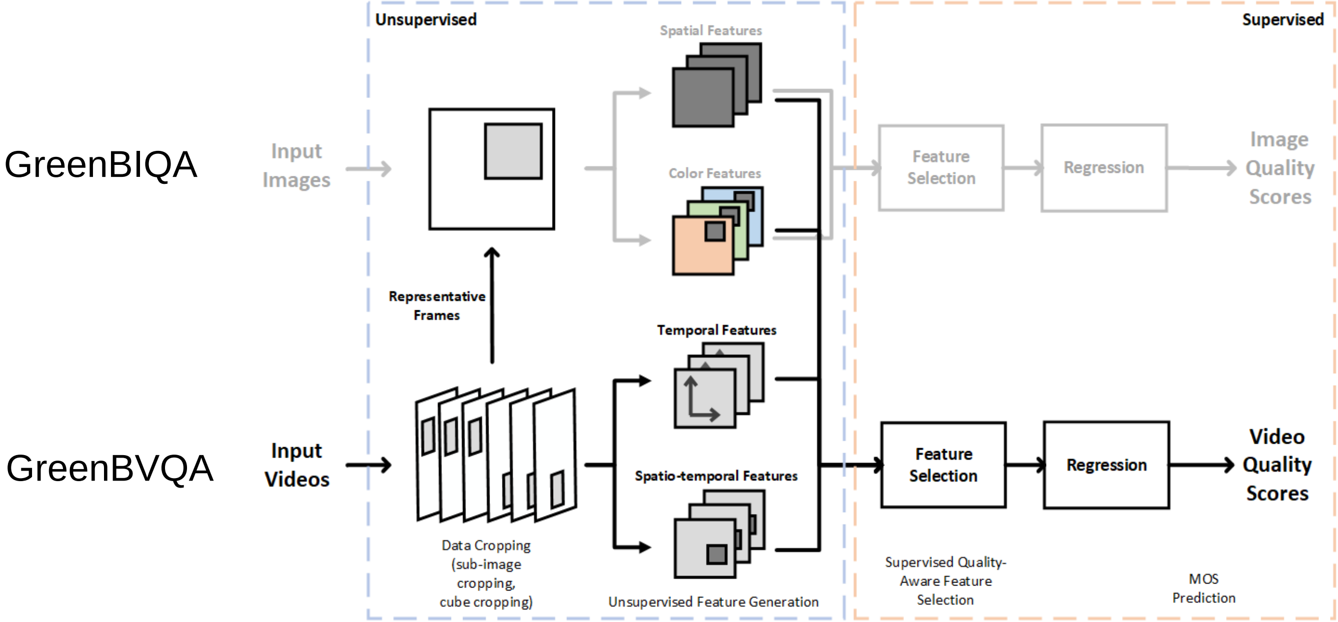
MOS: 4.02



MOS: 1.55

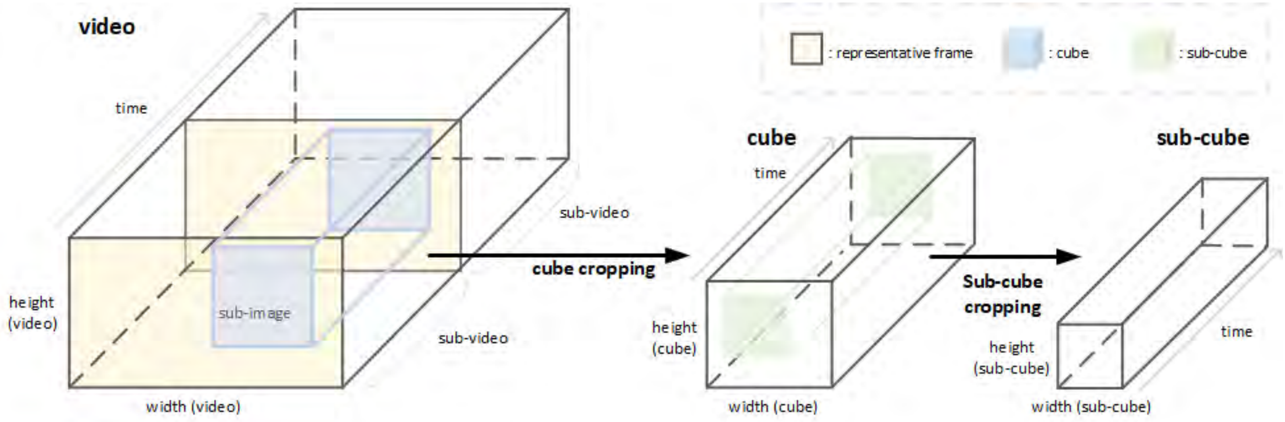
GreenBVQA - Pipeline

- Combine GreenBIQA and GreenBVQA to a systematic pipeline



GreenBVQA - Video features

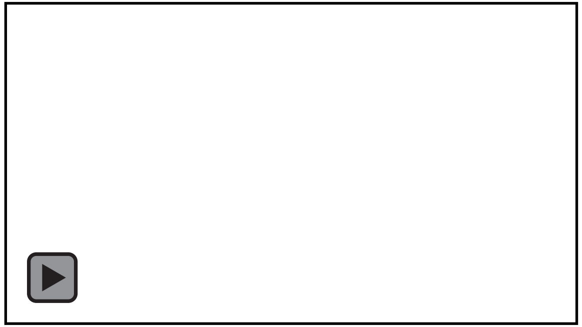
- Data cropping hierarchy for feature extraction
  - Frame->sub-image, video->sub-video, cube->sub-cube
  - Sub-image: spatial feature (2D-transform)
  - Cube: temporal, spatial-temporal feature (3D-transform)
  - Sub-cube: color feature (3D-transform)



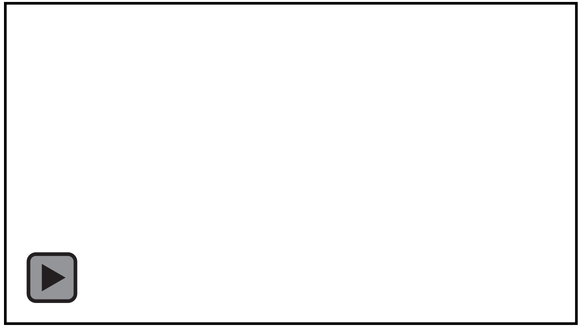
GreenBVQA – Prediction Results

MOS: 2.61
Predict: 2.68
Accurate

MOS: 3.60
Predict: 3.52
Accurate



MOS: 4.02
Predict: 2.97
Under-
estimate



MOS: 1.55
Predict: 2.30
Over-
estimate

GreenBVQA - Performance Benchmarking

- Model complexity comparison, where the reported SROCC and PLCC performance numbers are against the KoNViD-1k dataset

Model	SROCC↑	PLCC↑	Model Size (MB)↓	FLOPs↓
VSFA [37]	0.794	0.798	100.2 (15.8×)	20T (1250×)
QSA-VQM [39]	0.801	0.802	196 (30.8×)	40T (2500×)
Mirko *et al.* [47]	0.772	0.784	42.3 (6.6×)	1.5T (94×)
CNN-TLVQM [40]	**0.814**	**0.817**	98 (15.4×)	21T (1312×)
GreenBVQA(Ours)	0.776	0.779	**6.36 (1×)**	**16G (1×)**

Model Performance (Correlations with MOS)

Model Memory Efficiency

Computational Complexity

Takeaway

- Objective quality assessment for images and videos is essential in RTC
- There is no reference available in UGC and conversational video
- Our proposed solution, GreenBIQA and GreenBVQA, can achieve tier-one performance with ~50x smaller model size and ~500x less computational complexity as compared to SOTA deep learning methods
- Weakly supervised learning is one of the future research directions

Reference (BIQA)

[1] Bosse, S., Maniry, D., M¨uller, K.-R., Wiegand, T., and Samek, W. (2017). Deep neural networks for no-reference and full-reference image quality assessment. IEEE Transactions on image processing, 27(1):206–219.

[2] Kim, J. and Lee, S. (2016). Fully deep blind image quality predictor. IEEE Journal of selected topics in signal processing, 11(1):206–220.

[3] Mittal, A., Moorthy, A. K., and Bovik, A. C. (2012a). No-reference image quality assessment in the spatial domain. IEEE Transactions on image processing, 21(12):4695–4708.

[4] Mittal, A., Soundararajan, R., and Bovik, A. C. (2012b). Making a “completely blind” image quality analyzer. IEEE Signal processing letters, 20(3):209–212.

[5] Su, S., Yan, Q., Zhu, Y., Zhang, C., Ge, X., Sun, J., and Zhang, Y. (2020). Blindly assess image quality in the wild guided by a self-adaptive hyper network. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 3667–3676.

[6] Talebi, Hossein, and Peyman Milanfar. "NIMA: Neural image assessment." IEEE transactions on image processing 27.8 (2018): 3998-4011.

[7] Xu, J., Ye, P., Li, Q., Du, H., Liu, Y., and Doermann, D. (2016). Blind image quality assessment based on high order statistics aggregation. IEEE Transactions on Image Processing, 25(9):4444–4457.

[8] Ye, P., Kumar, J., Kang, L., and Doermann, D. (2012). Unsupervised feature learning framework for no-reference image quality assessment. In 2012 IEEE conference on computer vision and pattern recognition, pages 1098–1105. IEEE.

[9] Zeng, H., Zhang, L., and Bovik, A. C. (2018). Blind image quality assessment with a probabilistic quality representation. In 2018 25th IEEE International Conference on Image Processing (ICIP), pages 609–613. IEEE.

[10] Zhang, W., Ma, K., Yan, J., Deng, D., and Wang, Z. (2018). Blind image quality assessment using a deep bilinear convolutional neural network. IEEE Transactions on Circuits and Systems for Video Technology, 30(1):36–47. 5

Reference (BVQA)

[1] He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 770–778.
[2] Li, B., Zhang, W., Tian, M., Zhai, G., and Wang, X. (2022). Blindly assess quality of in-the-wild videos via quality-aware pre-training and motion perception. IEEE Transactions on Circuits and Systems for Video Technology.
[3] Li, D., Jiang, T., and Jiang, M. (2019). Quality assessment of in-the-wild videos. In Proceedings of the 27th ACM International Conference on Multimedia, pages 2351–2359.
[4] Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.
[5] Tu, Z., Yu, X., Wang, Y., Birkbeck, N., Adsumilli, B., and Bovik, A. C. (2021). Rapique: Rapid and accurate video quality prediction of user generated content. IEEE Open Journal of Signal Processing, 2:425–440.
[6] Ying, Z., Mandal, M., Ghadiyaram, D., and Bovik, A. (2021). Patch-vq:'patching up'the video quality problem. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 14019–14029.

Thank You.

We are happy to answer any questions.

GreenBVQA – Performance on individual datasets

TABLE VI
COMPARISON OF THE PLCC AND SROCC PERFORMANCE OF 10 BENCHMARKING METHODS AGAINST THREE VQA DATASETS.

	CVD2014		LIVE-VQC		KoNViD-1k		Average	
Model	SROCC↑	PLCC↑	SROCC↑	PLCC↑	SROCC↑	PLCC↑	SROCC↑	PLCC↑
NIQE [4]	0.475	0.607	0.593	0.631	0.539	0.551	0.535	0.596
BRISQUE [6]	0.790	0.804	0.593	0.624	0.649	0.651	0.677	0.654
CORNIA [7]	0.627	0.663	0.681	0.723	0.735	0.735	0.681	0.707
V-BLIINDS [13]	0.795	0.806	0.681	0.699	0.706	0.701	0.727	0.735
TLVQM [32]	0.802	0.823	0.783	0.785	0.763	0.765	0.782	0.791
VIDEVAL [54]	0.814	0.832	0.744	0.748	0.770	0.771	0.776	0.783
VSFA [37]	0.850	<u>0.859</u>	0.717	0.770	0.794	0.798	0.787	0.809
RAPIQUE [18]	0.807	0.823	0.741	0.761	0.788	<u>0.805</u>	0.778	0.796
QSA-VQM [39]	<u>0.850</u>	<u>0.859</u>	0.742	0.778	<u>0.801</u>	0.802	0.797	<u>0.813</u>
Mirko *et al.* [47]	0.834	0.848	0.742	0.780	0.772	0.784	0.782	0.804
CNN-TLVQM [40]	**0.852**	**0.868**	**0.811**	**0.828**	**0.814**	**0.817**	**0.825**	**0.837**
GreenBVQA(Ours)	0.835	0.854	<u>0.785</u>	<u>0.789</u>	0.776	0.779	<u>0.798</u>	0.807

GreenBVQA – Computation efficiency

- Three settings of videos (240frs@540p, 364frs@480p, 467frs@720p)

INFERENCE TIME COMPARISON IN SECONDS.

Model	240frs@540p	364frs@480p	467frs@720p
V-BLIINDS [13]	382.06	361.39	1391.00
QSA-VQM [39]	281.21	256.13	900.72
VSFA [37]	269.84	249.21	936.84
TLVQM [32]	50.73	46.32	136.89
NIQE [4]	45.65	41.97	155.90
BRISQUE [6]	12.69	12.34	41.22
Mirko *et al.* [47]	8.43	6.24	16.29
GreenBVQA	**3.22**	**4.88**	**6.26**

GreenBVQA – Computation efficiency

- Three settings of videos (240frs@540p, 364frs@480p, 467frs@720p)