Video compression dataset and benchmark of learning-based video-quality metrics

Anastasia Antsiferova, Sergey Lavrushkin,
Maksim Smirnov, Alexander Gushchin,
Dmitriy Vatolin, Dmitriy Kulikov

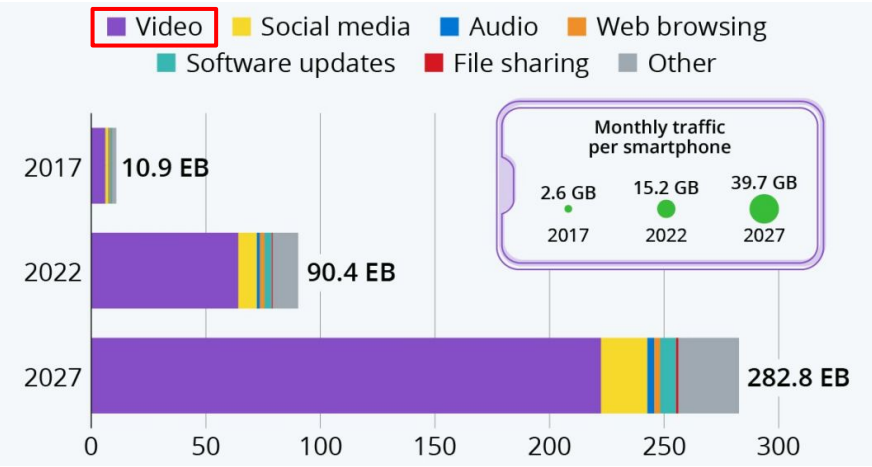
CS MSU Graphics&Media Lab

Introduction

The importance of video quality assessment

- Video takes >80% of world Internet traffic
- New video compression standards provide 3 times lower bitrate than older standards, thanks to new algorithms including neural networks

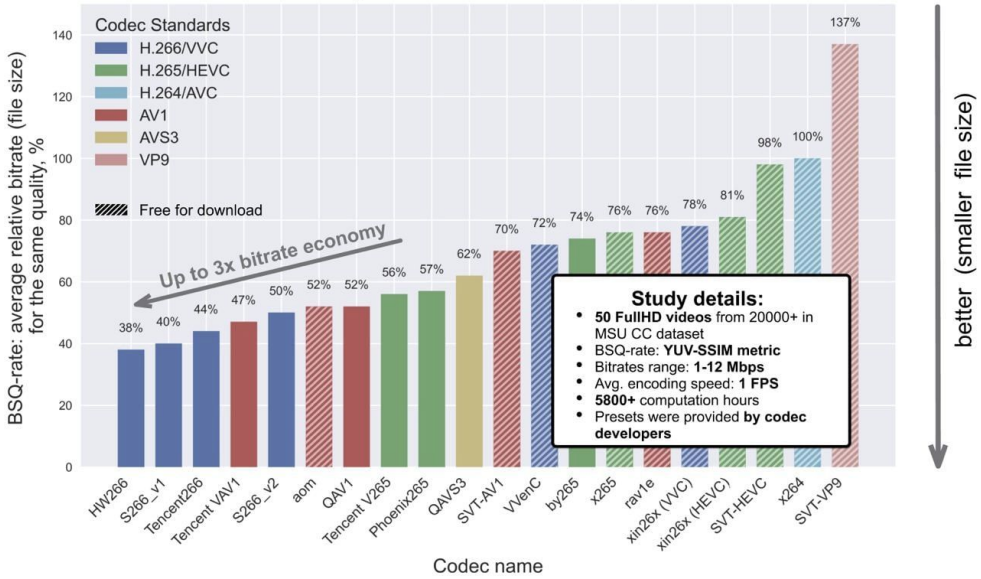
Estimated global mobile data traffic (exabytes/month)



Growth in mobile video traffic consumption (Ericsson Mobility Report 2022)

VVC codecs superiority in MSU Codec Comparison 2021

https://compression.ru/video/codec_comparison/2021



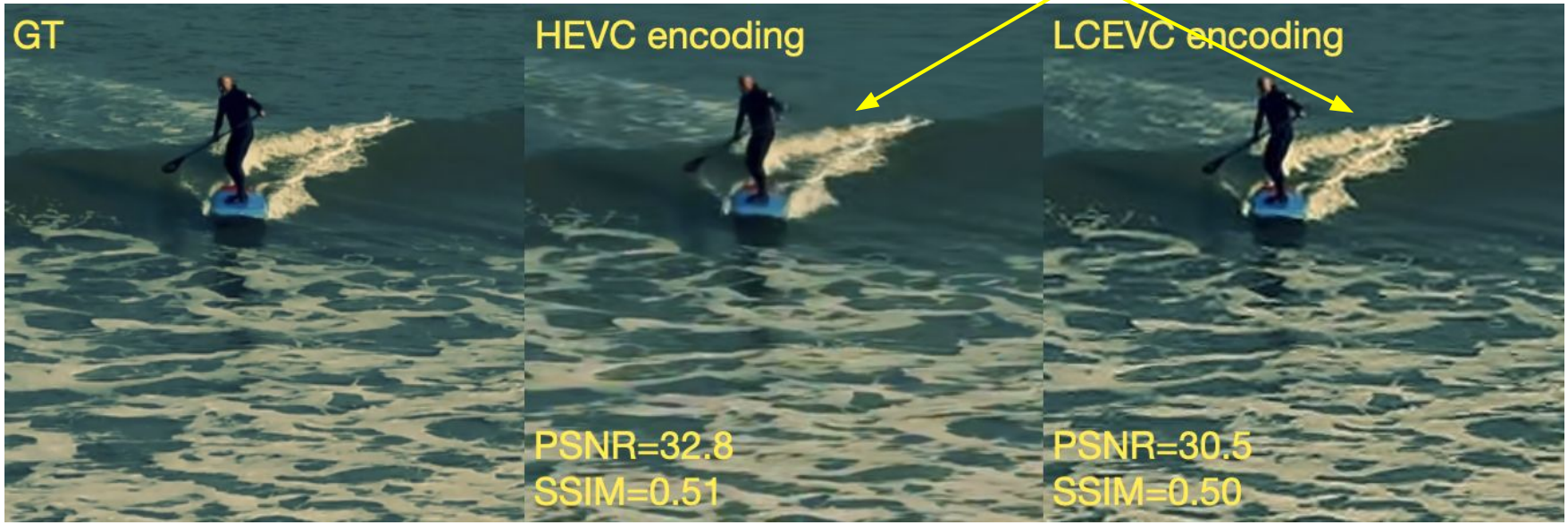
Compression rate provided by different encoding standards

Introduction

Drawbacks of existing video quality metrics research

Existing comparisons of quality metrics use videos encoded by H.264 and H.265 and do not analyse distortions arising from new compression standards

LCEVC employs super-resolution, which allows restoration of more details, however PSNR and SSIM scores are identical



Crops from video sequences encoded using x265 and lcevc_x265 relative to original (GT)

Introduction
Our contributions

Our goal is to evaluate new image- and video-quality metrics using a large dataset representing diverse compression artifacts

- We propose a new dataset of 2,486 compressed videos and subjective scores collected using a crowdsourced comparison with nearly 11,000 participants
- We also present a new benchmark based on that dataset, which we divided into open and hidden parts

Proposed dataset

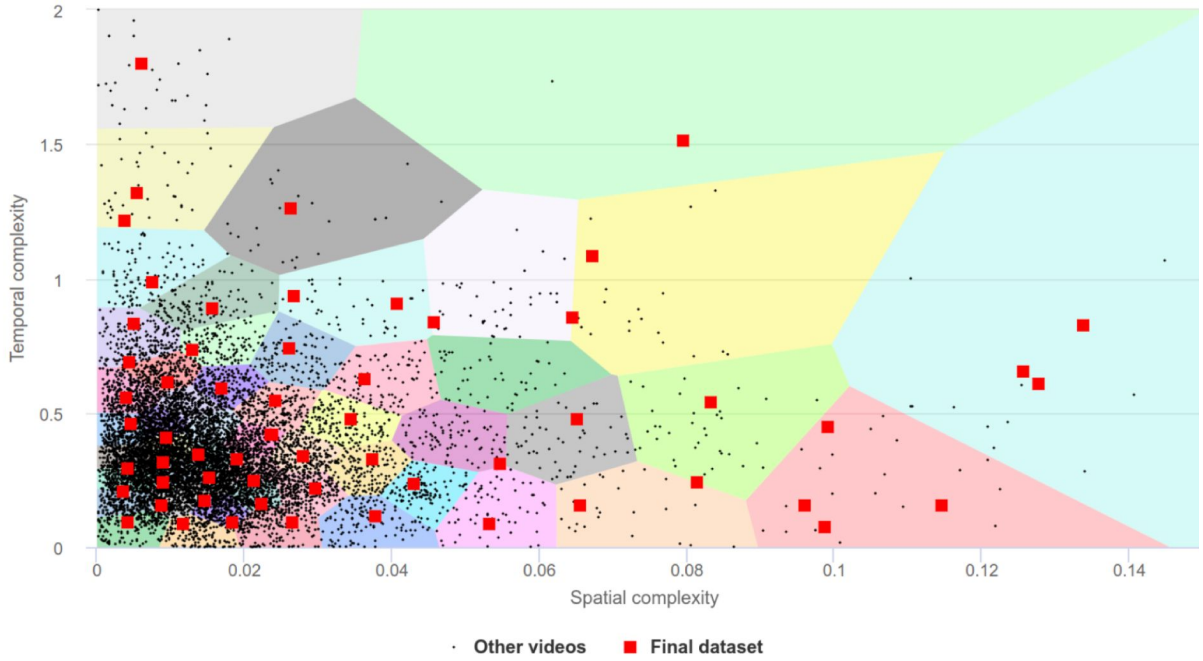
Existing subjective datasets with compression

Dataset	Original videos	Average duration (s)	Distorted videos	Distortion	Subjective framework	Subjects	Answers
MCL-JCV (2016) [H.Wang et al.]	30	5	1,560	Compression	In-lab	150	78K
VideoSet (2017) [H. Wang et al.]	220	5	45,760	Compression	In-lab	800	-
UGC-VIDEO (2020) [Y. Li et al.]	50	>10	550	Compression	In-lab	30	16.5K
CVD-2014 [M. Nuutinen el al.]	5	10-25	234	In-capture	In-lab	210	-
LIVE-Qualcomm [D. Ghadiyaram et al.]	54	15	208	In-capture	In-lab	39	8.1K
GamingVideoSET [N. Barman et al.]	24	30	576	Compression	In-lab	25	-
KUGVD (2019) [N. Barman et al.]	6	30	144	Compression	In-lab	17	-
KoNViD-1k (2017) [V. Hosu et al.]	1,200	8	1,200	In-the-wild	Crowdsource	642	205K
LIVE-VQC (2018) [Z. Sinno et al.]	585	10	585	In-the-wild	Crowdsource	4,776	205K
YouTube-UGC (2019) [Y. Wang et al.]	1,500	20	1,500	In-the-wild	Crowdsource	>8,000	600K
LSVQ (2020) [Z. Ying et al.]	39,075	5-12	39,075	In-the-wild	Crowdsource	6,284	5M
Proposed dataset (2022)	36	10, 15	2,486	Compression (83 codecs)	Crowdsource	10,800	766K

Proposed dataset

Video selection

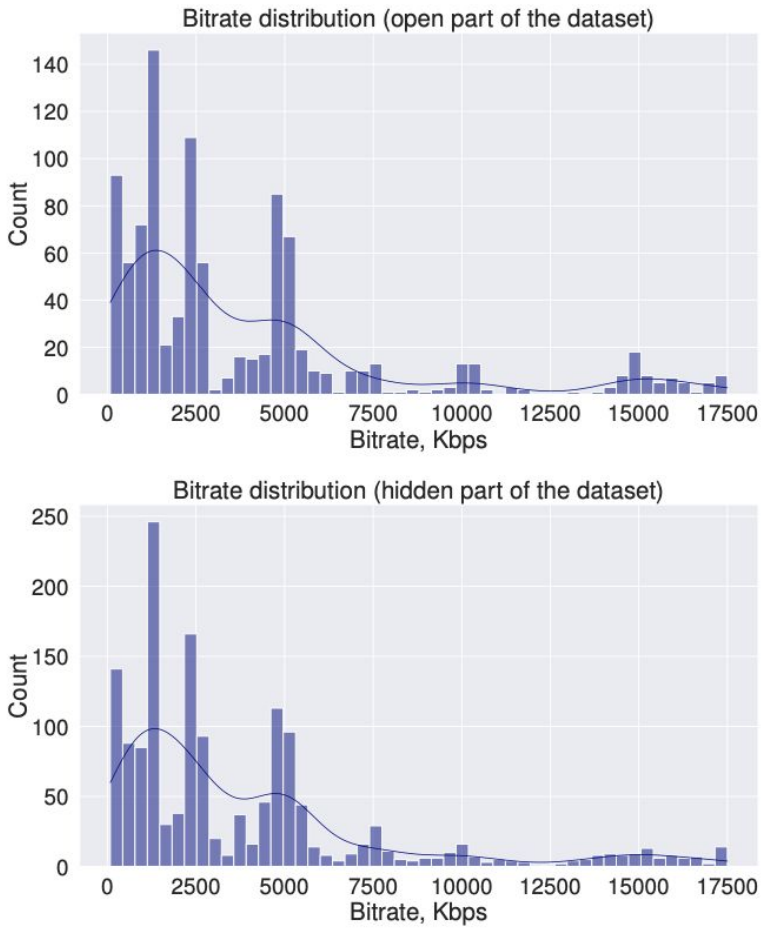
- We downloaded 18,000 videos with high bitrate and resolution from vimeo.com website, xiph.org and Youtube UGC collections
- Clustered them by spatial and temporal complexity
- 36 selected videos have different FPS and 10s duration



Selected videos and spatio-temporal clustering of video collection

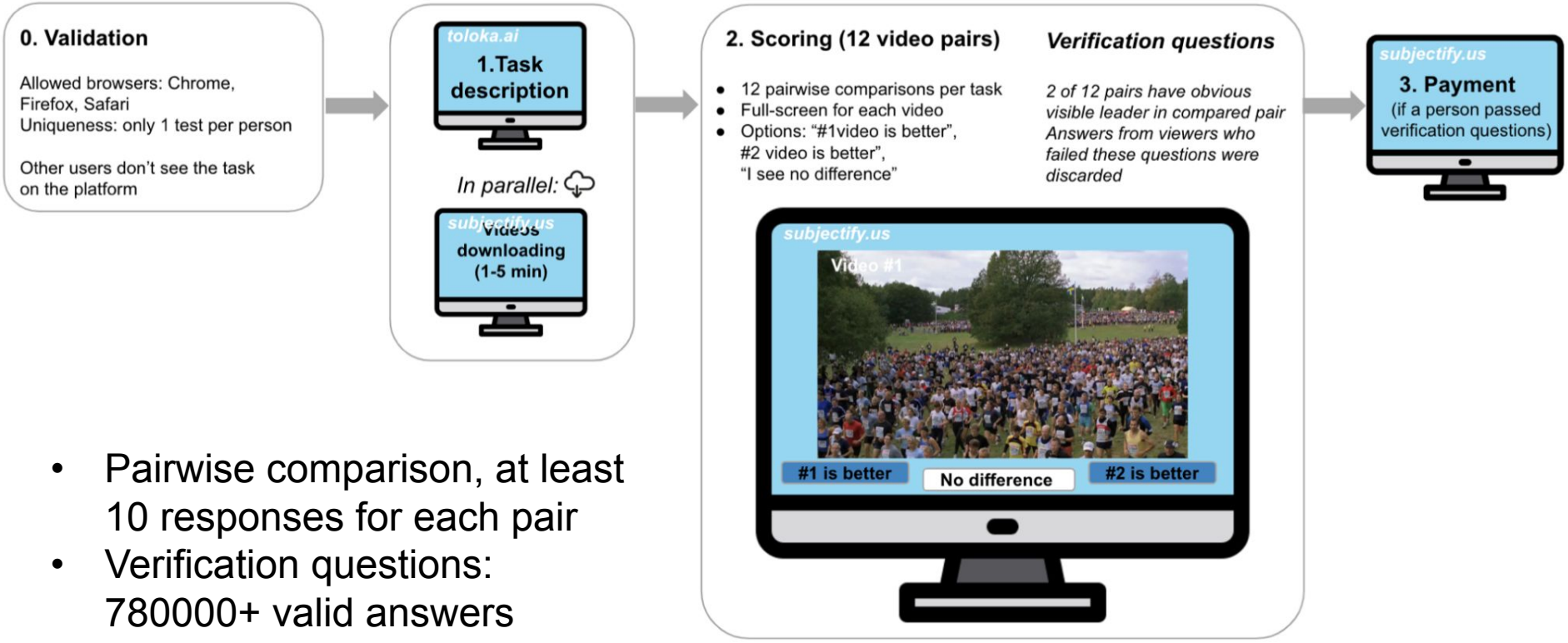
Proposed dataset
Compression

- **83 codecs**
  (different codecs versions and encoding presets)
- **7 compression standards:**
  H.264/AVC, H.265/HEVC, AV1, H.266/VVC, VP9, etc.
- **3 target bitrates:** 1,000 kbps, 2,000 kbps, and 4,000 kbps
- Dataset is divided into open (32 codecs) and hidden part



Proposed dataset
Subjective comparisons

Subjective assessment was conducted using *Subjectify.us*



Subjective assessment scheme

- Pairwise comparison, at least 10 responses for each pair
- Verification questions: 780000+ valid answers
- 10800 unique participants

Dataset download link

Open part of the dataset is publicly available:

https://calypso.gml-team.ru:5001/sharing/lxSWi6vtg

Password: c943=R3/tJwVV%P%

Proposed benchmark

Other benchmarks

The majority of image/videos quality metrics comparisons were published in papers and are not updated

Benchmark	# videos	# methods	# subjects	Distortion
Z. Sinno and A. Bovik (2018)	585	4	4,776	In-the-wild videos (80 mobile cameras, 18 resolutions)
Y. Li et al. (2020)	550	15	28	Compression (H.264, H.265, QP: 22, 27, 32, 37, 42)
UGC-VQA http://ugcvqa.com/ (2021)	3,108 (LIVE-VQC, YouTube-UGC, KoNViD-1k)	23	>13,000	Compression, transmission
Proposed benchmark (2022)	2,486	27	10,800	Compression (H.264, H.265, AV1, VVC, etc)

Proposed benchmark

Participants and categories

- Compared 12 full-reference and 15 no-reference metrics
- Results given for all video set and subsets:
  - By compression standard (H.265/HEVC, H.266/VVC, AV1)
  - By bitrate range (<1,000 kbps and >6,000 kbps)
  - By video types (UGC, Shaking, Sports, Nature, Gaming & Animation)
- Calculated correlation coefficients for metrics scores with subjective scores

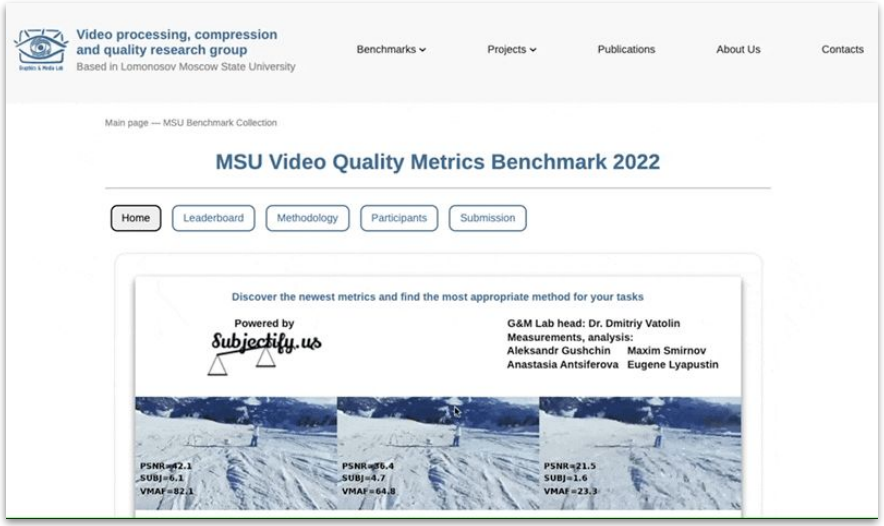
Name	Iqa/Vqa	Type	Added By	Year
AVQT	VQA	FR	MSU	2021
DISTS [0]	IQA	FR	MSU	2020
FovVideoVDP [1]	VQA	FR	MSU	2021
GVSP-UGCVQA-FR [2]	VQA	FR	Submit	2021
GVSP-UGCVQA-NR [3]	VQA	NR	Submit	2021
KonCept512 [4]	IQA	NR	MSU	2020
LINEARITY [5]	IQA	NR	MSU	2020
LPIPS [6]	IQA	FR	MSU	2018
MDTVSFA [7]	VQA	NR	MSU	2021
MEON [8]	IQA	NR	MSU	2017
MS-SSIM [9]	IQA	FR	MSU	2004
NIMA [10]	IQA	NR	MSU	2018
NIQE [11]	IQA	NR	MSU	2012
PAQ-2-PIQ [12]	IQA	NR	MSU	2020
PSNR	IQA	FR	MSU	–
SPAQ [13]	IQA	NR	MSU	2020
SSIM [14]	IQA	FR	MSU	2003
ST-GREED [15]	VQA	FR	MSU	2021
Tencent DVQA	VQA	FR	MSU	2020
TLVQM [16]	VQA	NR	MSU	2019
UNIQUE [17]	IQA	NR	Submit	2020
VIDEVAL [18]	VQA	NR	MSU	2021
VMAF [19]	VQA	FR	MSU	2017
VQM	VQA	FR	MSU	–
VSFA [20]	VQA	NR	MSU	2019

List of compared metrics

Proposed benchmark
Project page

Benchmark results and methodology page:

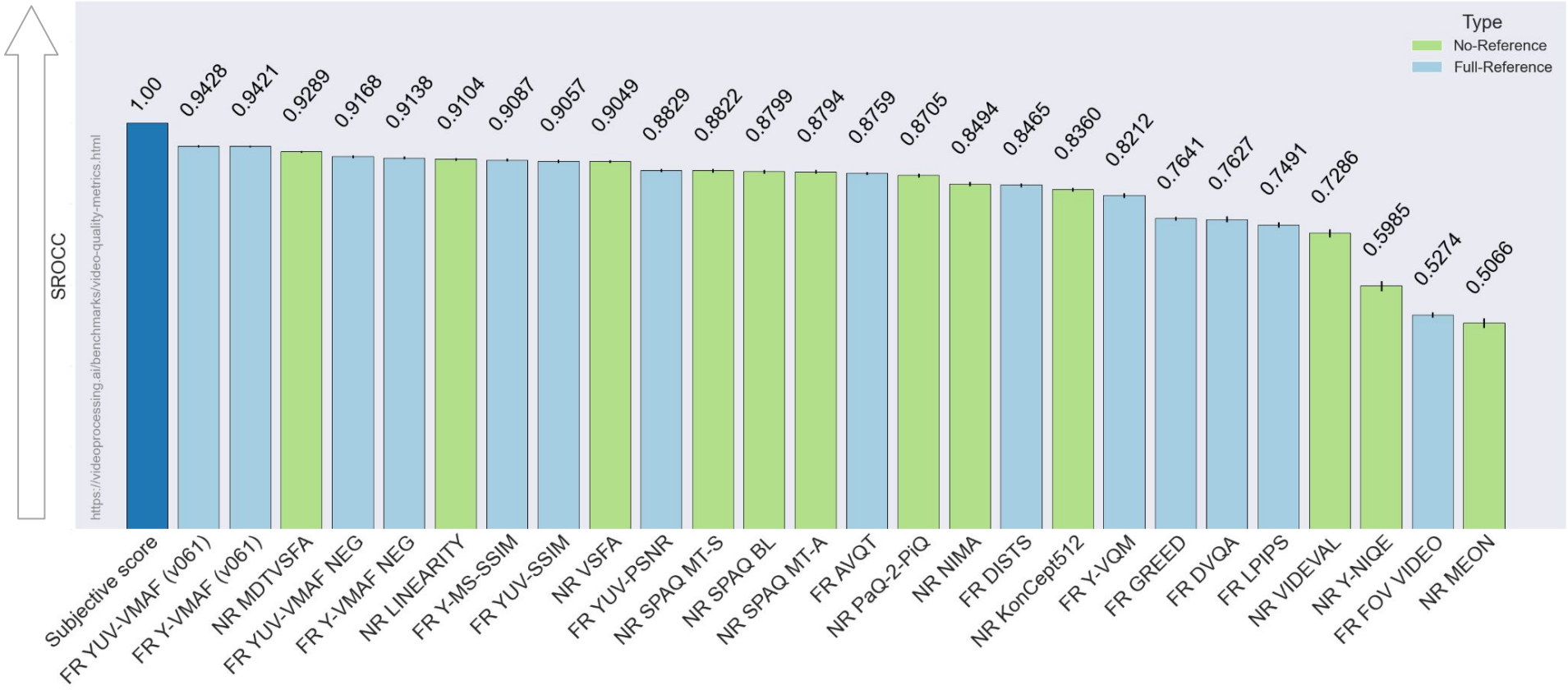
https://videoprocessing.ai/benchmarks/video-quality-metrics.html



Results

Spearman CC on full dataset

- New no-reference methods catch up with full-reference ones
- VMAF and MDTVSFA showed the best correlations



Results

SROCC by different categories

	Low Bitrate	High Bitrate	H.265	AV1	VVC	UGC	Shaking	Sports	Nature	Gaming & Animation
No-ref. metrics	MDTVSFA (0.943)	MDTVSFA (0.560)	MDTVSFA (0.945)	MDTVSFA (0.932)	Linearity (0.919)	MDTVSFA (0.912)	MDTVSFA (0.897)	Linearity (0.931)	MDTVSFA (0.917)	VSFA (0.975)
	NIMA (0.904)	VSFA (0.517)	Linearity (0.932)	VSFA (0.914)	SPAQ MT-S (0.908)	Linearity (0.893)	Linearity (0.864)	MDTVSFA (0.924)	Linearity (0.904)	MDTVSFA (0.971)
	Linearity (0.900)	Linearity (0.470)	VSFA (0.927)	Linearity (0.906)	SPAQ MT-A (0.894)	VSFA (0.852)	NIMA (0.849)	VSFA (0.911)	VSFA (0.873)	Linearity (0.964)
Full-ref. metrics	YUV-VMAF (0.952)	Y-VMAF (0.453)	YUV-VMAF (0.946)	YUV-VMAF (0.910)	YUV-SSIM (0.912)	Y-VMAF (0.946)	Y-VMAF (0.891)	YUV-VMAF (0.961)	YUV-VMAF (0.944)	YUV-VMAF (0.972)
	Y-MS-SSIM (0.952)	DISTS (0.417)	Y-VMAF (0.940)	Y-VMAF (0.905)	Y-MS-SSIM (0.905)	YUV-VMAF (0.942)	YUV-VMAF (0.879)	Y-MAF (0.959)	Y-VMAF (0.942)	YUV-VMAF NEG (0.968)
	Y-VMAF NEG (0.946)	YUV-SSIM (0.302)	YUV-VMAF NEG (0.920)	YUV-SSIM (0.869)	YUV-PSNR (0.900)	AVQT (0.919)	YUV-VMAF NEG (0.849)	YUV-SSIM (0.952)	YUV-VMAF NEG (0.913)	Y-VMAF (0.967)

Conclusion

- A new dataset with 83 compression types is proposed, open part is publicly available
- A new benchmark of 26 full-reference and no-reference image/video quality metrics is published. Since publication, we received 4 new submissions
- Positive feedback from Huawei, Tencent, Yandex, Google (YouTube Media Algorithms) and others
- We plan to expand the dataset with new compression standards, codecs and content

MSU benchmarks

GRAPHICS & MEDIA LAB
VIDEO GROUP

https://videoprocessing.ai/benchmarks/

News

- **09.10.2022** [MSU Full-Reference Video Quality Metric Benchmark] Added new algorithms and Cite Us section
- **09.10.2022** [MSU No-Reference Video Quality Metric Benchmark] Added new algorithms and Cite Us section
- **28.08.2022** [MSU Video Upscalers Benchmark 2022: Quality Enhancement] Release of the benchmark
- **08.06.2022** [MSU HDR Video Reconstruction Benchmark 2022] Alpha-version Release
- **06.04.2022** [MSU Video Super Resolution Benchmark: Detail Restoration] Added 8 new algorithms and LPIPS metric
- **16.03.2022** [MSU Video Super Resolution Benchmark: Detail Restoration] Preprint of our paper "Towards True Detail Restoration for Super-Resolution: A Benchmark and a Quality Metric" was released on arXiv
- **15.11.2021** [MSU Video Super Resolution Benchmark: Detail Restoration] Our paper "ERQA: Edge-restoration Quality Assessment for Video Super-Resolution" was accepted to VISAPP 2022
- **09.11.2021** [MSU Video Upscalers Benchmark 2022: Quality Enhancement] Alpha-version Release
- **26.10.2021** [MSU Super-Resolution for Video Compression Benchmark 2021] Updated the Methodology
- **22.10.2021** [MSU Video Alignment and Retrieval Benchmark] Public beta-version Release
- **12.10.2021** [MSU Super-Resolution for Video Compression Benchmark 2021] Published October Report. Added 2 new videos to the dataset. Updated Charts section and Visualizations
- **06.10.2021** [MSU Video Alignment and Retrieval Benchmark] Alpha-version Release
- **01.10.2021** [MSU Mobile Video Codecs Benchmark 2021] Beta-version Release
- **14.09.2021** [MSU Super-Resolution for Video Compression Benchmark 2021] Public beta-version Release
- **31.08.2021** [MSU Super-Resolution for Video Compression Benchmark 2021] Alpha-version Release
- **26.04.2021** [MSU Video Super Resolution Benchmark: Detail Restoration] Beta-version Release
- **05.05.2021** [MSU Shot Boundary Detection Benchmark 2020] Main Release