The Frankenstone toolbox for video quality analysis of user-generated content

Steve Göring, Alexander Raake

Audiovisual Technology Group; Technische Universität Ilmenau, Germany
Email: [steve.goering, alexander.raake]@tu-ilmenau.de

Code & Data: `http://git.avt-imt.de/frankenstone`

Introduction

- video streaming: major part of the overall bandwidth of the internet [2]
- popular user-generated content (UGC): YouTube, TikTok, Facebook, ...
- **challenges:** video quality prediction and tuning of video encoders for UGC
- UGC datasets:
  - YouTube UGC Dataset (YTUGC) [13]
  - KoNViD-1k [9], LIVE Wild Compressed Video Quality Database [17]
- UGC: linking visual quality and aesthetics
- UGC prediction models:
  - dover model [14], fast(er)VQA [15], Q-Align [16] ...
- **missing**: unified common toolbox for various models/features
  - → the Frankenstone toolbox

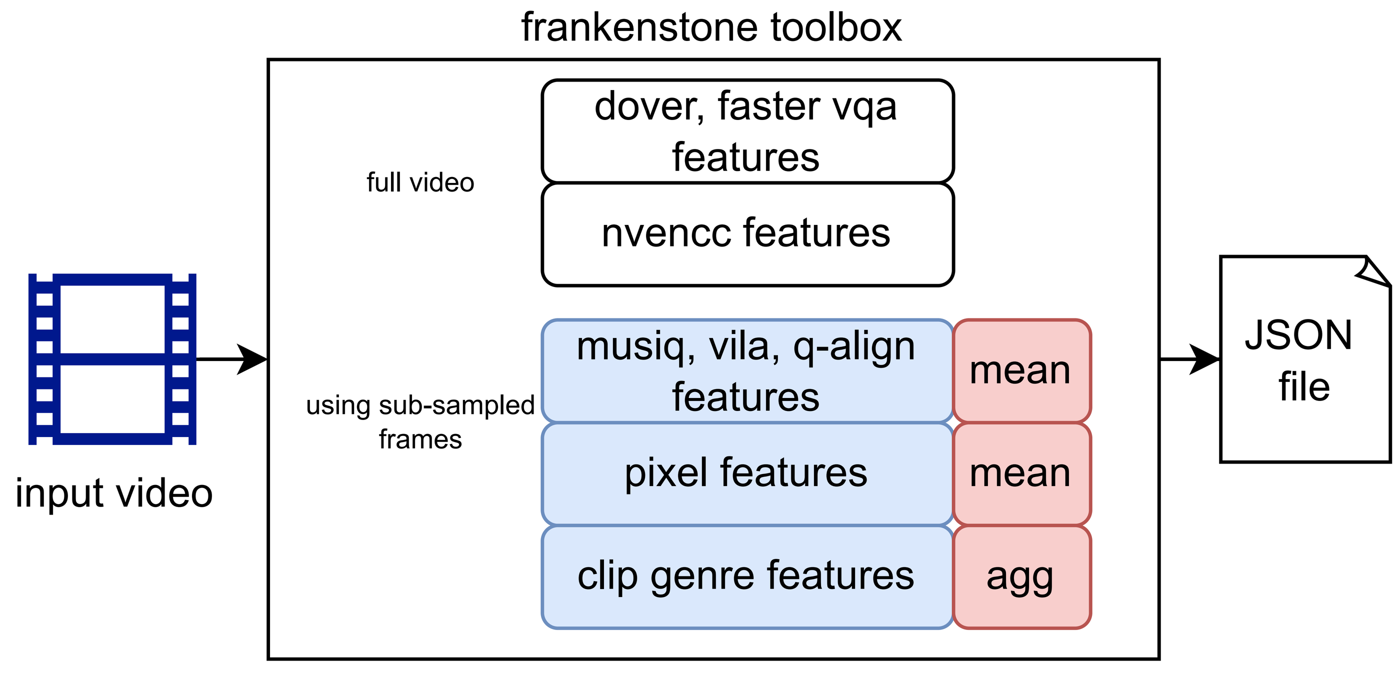

Figure 1: Overview of the Frankenstone toolbox.

Overview of the Frankenstone toolbox

- overview of toolbox in figure 1
  - input: video (no reference)
  - parallel computation of features/models; frame subsampling
  - aggregation of results
  - store as JSON report
- frame subsampling:
  - dover, faster vqa: internal sub-sampling,
  - NVENCC: no sub-sampling
  - other features: sub-sampling: first frame of each second of the video
- requirements
  - Python 3.11 with Tensorflow, PyTorch;
  - recent GPU with 24GB memory, e.g., NVIDIA GeForce RTX 3090 Ti
- implementation: open source
- parallel execution:
  - fully utilization of GPU
  - save computation time
- toolbox; extensions , modifications, proof of concept

Feature/model groups

- NVENCC [4]: H.265 hardware encoding (default parameters)
  - video complexity measure, similar to [5]
  - height, width, aspect ratio, I/P/B frame ratios, average I/P/B frame QP values, SSIM/PSNR to the input video, and bitrate
- Dover [14] and FasterVQA [15]:
  - Dover: fused score of aesthetics and quality + both individual scores
  - FasterVQA: overall and raw scores
- Pixel Features: re-implementation of some features from [7]
  - variant of SI/TI (no lumminance conversion), colorfulness [8], average lumminance
  - sharpness indicator feature (MSE to blurried frame)
  - NIMA appeal and quality scores [12] with TF-lite, TI-first, SSIM-first, SSIM-pair
  - all for full-frame and center-cropped variant (50% inner center crop), cmp. [6, 7]
- Musiq [10], Vila [11], and Q-Align [16]:
  - Vila: TFHub, pre-trained model; Musiq + Q-Align: IQA-PyTorch
- CLIP Genre:
  - per frame similarity with Open-CLIP [1] to 16 text prompts, e.g., "animated photo", "portrait photo" (per frame max matching)
  - how often different genres, most often detected genre

Evaluation

- runtime of the tool: table 1
  - each feature group
  - all (parallel) vs. sequential
- features vs. human ratings: figure 3, 2
  - test subset of YTUGC [13] (125 videos: 360p, 480p, 720p, 1080p, 2160p)
  - 16 categories (animated to virtual reality videos)

Table 1: P-time for the feature groups and all.

feature group	mean_time [s]	std_time	% compared to all
faster vqa	6.44	0.01	12
clip genre	7.33	0.06	14
dover	9.47	0.03	17
nvencc	10.32	0.00	19
pxl	14.90	0.07	27
musiq	15.75	0.79	29
vila	22.74	0.09	42
q-align	27.80	0.06	51
all	54.48	0.31	100
sequential (all)	80.27	0.46	147

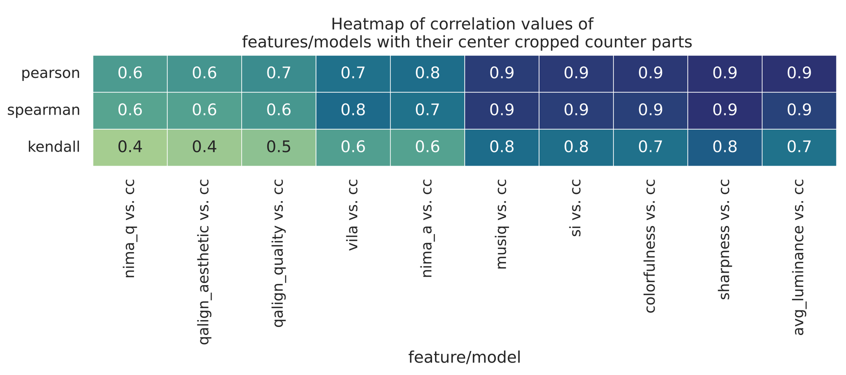

Figure 2: Heatmap of correlation values for all feature/model values compared to their center cropped variants.

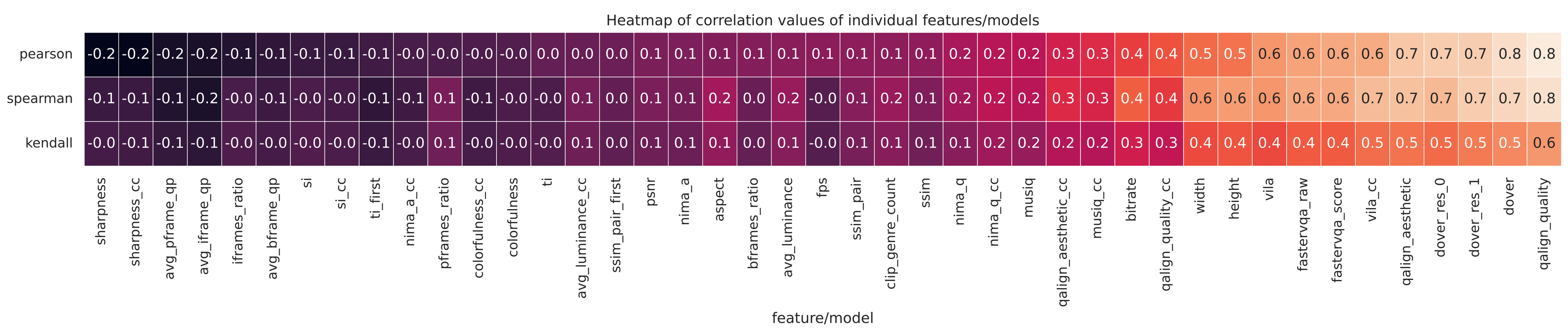

Figure 3: Heatmap of correlation values for all included feature/model values.

Conclusion and Future Work

- a unified toolbox – Frankenstone– for UGC/video quality:
  - video quality models, meta-data, and signal-based feature extraction to analyze UGC
  - GPU based, utilization
- proof-of-concept: extensions possible
- evaluation: runtime and performance
- building block for video quality research, see [3]

References

[1] M. Cherti et al. "Reproducible scaling laws for contrastive language-image learning". In: *CVF*. 2023.
[2] Cisco. *Cisco Visual Networking Index: Forecast and Trends, 2017–2022*. 2022.
[3] M. V. Conde et al. *AIS 2024 challenge on video quality assessment of UGC: Methods and results*. 2024.
[4] *Github NVEnc*. URL: https://github.com/rigaya/NVEnc (visited on 04/09/2024).
[5] S. Göring et al. *AVT-VIBE – Overview of two Models for the ICIP 2024 Grand Challenge On Video Complexity*.
[6] S. Göring et al. *cencro – Speedup of Video Quality Calculation using Center Cropping*. 2019.
[7] S. Göring et al. *Modular Framework and Instances of Pixel-based Video Quality Models for UHD-1/4K*. 2021.
[8] D. Hasler et al. "Measuring colorfulness in natural images". In: *HVEI*. 2003.
[9] V. Hosu et al. "The Konstanz natural video database (KoNViD-1k)". In: *QoMEX*. 2017.
[10] J. Ke et al. "Musiq: Multi-scale image quality transformer". In: *CVF*. 2021.
[11] J. Ke et al. *Vila: Learning image aesthetics from user comments with vision-language pretraining*. 2023.
[12] C. Lennan et al. *Image Quality Assessment*. 2018.
[13] Y. Wang et al. "YouTube UGC dataset for video compression research". In: *MMSP*. 2019.
[14] H. Wu et al. "Exploring Video Quality Ass. on UGC from Aesthetic and Technical Perspectives". In: *ICCV*. 2023.
[15] H. Wu et al. *Neighbourhood Representative Sampling for Efficient End-to-end Video Quality Assessment*. 2022.
[16] H. Wu et al. "Q-align: Teaching lmms for visual scoring via discrete text-defined levels". In: *arXiv* (2023).
[17] X. Yu et al. *Predicting the quality of compressed videos with pre-existing distortions*. 2021.

Acknowledgment

The authors want to thank the "AG Wissenschaftliches Rechnen" of the TU Ilmenau for providing computing resources.

TECHNISCHE UNIVERSITÄT ILMENAU