

Analyzing and Evaluating Image and Video Quality with Machine Learning

Audiovisual Technology Group (AVT), Technische Universität Ilmenau

Steve Göring



April 25, 2024

Introduction – About me

- ▶ 2008 – 2013: master/bachelor computer science at TU Ilmenau:
 - databases, network, theo-inf, math
- ▶ 2013 – 2014: Bauhaus Uni Weimar:
 - Big Data Analytics,
 - Hadoop/MapReduce,
 - Information Retrieval
- ▶ 2016 – 2020: Phd at AVT:
 - video quality in HAS streaming,
 - image appeal/quality
- ▶ 2020 – now: Post-Doc
 - AI-generated content topics



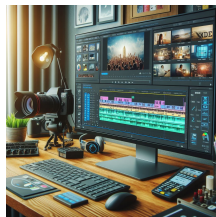
- ▶ Github: <https://github.com/stg7>
- ▶ Page: <https://stg7.github.io/>
- ▶ Email: steve.goering@tu-ilmenau.de

General Topic Scope



- ▶ mobile phones, camera/screen technology, more content + consumption
- ▶ **higher resolutions** (4K/8K screens, recordings); 2D content, internet:
 - **higher demand for video streaming**, up to 80% video streaming
 - **increase of uploaded images**, e.g., up to 95M per day for Instagram
 - internet bandwidth not necessarily adapting to trends
- ▶ new image/video compression methods

Video/Image Quality Models in the context of HAS streaming



- ▶ development of video/image quality models for higher resolutions
 - e.g. video UHD-1, images 4K resolution, gaming content
 - bitstream models: ITU-T P.1203¹, P.1204.3², **pixel/hybrid models**³, DNNs
- ▶ main focus in Phd thesis⁴: machine learning, computer vision

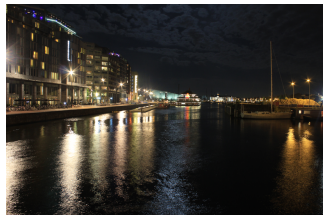
¹Robitza et al. "HTTP Adaptive Streaming QoE Estimation with ITU-T Rec. P.1203 – Open Databases and Software."

²Rao et al. "Bitstream-based Model Standard for 4K/UHD: ITU-T P.1204.3 – Model Details, Evaluation, Analysis and Open Source Implementation."

³Göring et al. "Modular Framework and Instances of Pixel-based Video Quality Models for UHD-1/4K."

⁴Göring. "Data-Driven Visual Quality Estimation Using Machine Learning." <https://doi.org/10.22032/dbt.52210>

Image Appeal Prediction



- ▶ evaluation of image appeal in a social media context⁵
- ▶ subjective testing lab vs. crowd, with likes|views /without
- ▶ prediction models using DNN, ML with signal, social, dnn-based, other features
 - reproducible research, **open science**, comparison to SoA datasets/work
- ▶ DFG project – “SoPhoAppeal”: Sept 2020 – Aug 2023

⁵Göring, Raake. “Image Appeal Revisited: Analysis, new Dataset and Prediction Models.” code: <https://git.avt-imt.de/sophoappeal>

AI generated Content

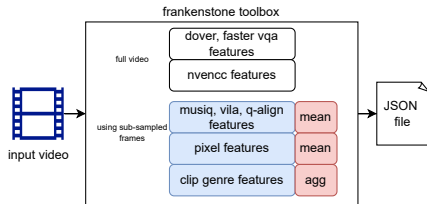


- ▶ AI content (text-to-X, enhancement): **new challenges**, e.g. distortions, realism
- ▶ AI-generated images⁶: focus text-to-image, realistic photos
 - evaluation of appeal, quality: subjective tests, objective models
- ▶ AI-based image upscaling ⁷
 - evaluation of different algorithms, detection (DNNs) and quality/appeal prediction
- ▶ future: **text-to-video/audio**, image/video compression

⁶Göring et al. "Analysis of Appeal for realistic AI-generated Photos." code: http://git.avt-imt.de/avt_ai_images

⁷Göring et al. "Appeal prediction for AI up-scaled Images." **under review**

Video Complexity and Quality of user-generated Content



```
1 #!/bin/bash
2 if [[ "$1" == "" ]]; then
3   echo "usage: ./est_sf_crf.sh <video>"
4   exit 0
5 fi
6 res=$(ffmpeg -y -i "$1" -filter:v "select='not(mod(n\,10))',setpts=N/FRAME_RATE/
7   ↳ TTB" -c:v libx264 -x264opts keyint=10min-keyint=10scenecut=-1-crf 28
8   ↳ -preset superfast -tune zerolatency -f mp4 /dev/null 2>&1 | grep "
9   ↳ libx264")
10 bitrate=$(echo "$res" | grep "kb/s" | sed "s|.*kb.s:||g")
11 echo "{\\"video\": \"$1\", \"estimated_bitrate_[kb/s]\": $bitrate}"
```

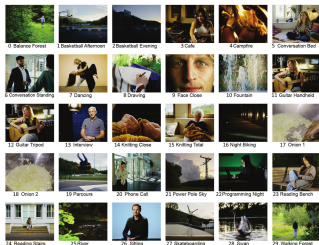
► Competitions using HPC/GPUs

- ITU Standardization
- video complexity (ICIP 2024 Challenge Challenge):
 - ▷ estimate target bitrate without full encoding⁸, **2nd and 3rd place**
- quality of user-generated content (Challenge at AIS2024 CVPR)⁹: **results pending**

⁸Göring, Rao. "AVT-VIBE – Overview of Two Models for the ICIP 2024 Grand Challenge.." code: http://git.avt-imt.de/avt_vibe

⁹Göring, Raake. "The Frankenstone toolbox for video quality analysis of user-generated content." code: <http://git.avt-imt.de/frankenstone>

Work and Projects at Audiovisual Technology Group (AVT)

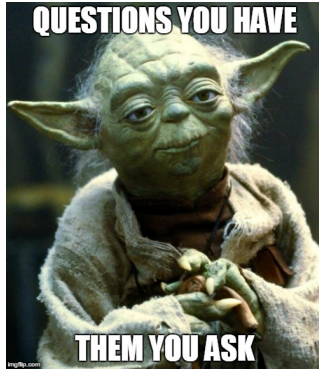


► Projects with focus

- on teleconferencing: 3D models, point clouds, avatars, immersion
- 8K Video, audio processing

► require: **large data**, **GPU processing**

Thank you for your attention



- ▶ prompt: an image with yoda saying “questions you have them ask you”
 - left: bing, middle: original, right: stable diffusion xl turbo