nofu – A Lightweight No-Reference Pixel Based Video Quality Model for Gaming Content.

Steve Göring, Rakesh Rao Ramachandra Rao, Alexander Raake;

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

June 6, 2019

TECHNISCHE UNIVERSITÄT
ILMENAU

Motivation – Gaming Streams

TECHNISCHE UNIVERSITÄT
ILMENAU

- beside classical video streams → gaming content:
  - e.g. Youtube Gaming, Twitch, ...
- gaming videos →
  - additional requirements /properties: *Zadtootaghaj et al.* [9]
  - live streaming, low delay, low stalling,
  - high video quality, cgi content, streaming technology
- focus on video quality of gaming streams

→ gaming qoe and gaming video quality

Gaming QoE and video quality

TECHNISCHE UNIVERSITÄT ILMENAU

- several influencing factors: *Möller, Schmidt, and Zadtootaghaj* [8]
  - video quality factors: content (cgi), encoding (fast),
  - interaction: delay, ...
- objective full-reference metrics: good results: *Barman et al.* [1, 2, 3]
  - VMAF best; problem: reference usually not available
- for live/adaptive encoding:
  - fast, accurate, no-reference quality estimation

$\rightarrow$ nofu

Gaming QoE and video quality

- several influencing factors: *Möller, Schmidt, and Zadtootaghaj* [8]
  - video quality factors: content (cgi), encoding (fast),
  - interaction: delay, . . .
- objective full-reference metrics: good results: *Barman et al.* [1, 2, 3]
  - VMAF best; problem: reference usually not available
- for live/adaptive encoding:
  - fast, accurate, no-reference quality estimation

$\rightarrow$ nofu

Gaming QoE and video quality

TECHNISCHE UNIVERSITÄT ILMENAU

- several influencing factors: *Möller, Schmidt, and Zadtootaghaj* [8]
  - video quality factors: content (cgi), encoding (fast),
  - interaction: delay, ...
- objective full-reference metrics: good results: *Barman et al.* [1, 2, 3]
  - VMAF best; problem: reference usually not available
- for live/adaptive encoding:
  - fast, accurate, no-reference quality estimation

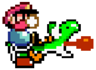
→ nofu

Gaming QoE and video quality

TECHNISCHE UNIVERSITÄT
ILMENAU

- several influencing factors: *Möller, Schmidt, and Zadtootaghaj* [8]
  - video quality factors: content (cgi), encoding (fast),
  - interaction: delay, ...
- objective full-reference metrics: good results: *Barman et al.* [1, 2, 3]
  - VMAF best; problem: reference usually not available
- for live/adaptive encoding:
  - fast, accurate, no-reference quality estimation

→ nofu



Gaming QoE and video quality

TECHNISCHE UNIVERSITÄT ILMENAU

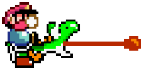
- several influencing factors: *Möller, Schmidt, and Zadtootaghaj* [8]
  - video quality factors: content (cgi), encoding (fast),
  - interaction: delay, ...
- objective full-reference metrics: good results: *Barman et al.* [1, 2, 3]
  - VMAF best; problem: reference usually not available
- for live/adaptive encoding:
  - fast, accurate, no-reference quality estimation

$\rightarrow$ nofu

Gaming QoE and video quality

- several influencing factors: *Möller, Schmidt, and Zadtootaghaj* [8]
  - video quality factors: content (cgi), encoding (fast),
  - interaction: delay, . . .
- objective full-reference metrics: good results: *Barman et al.* [1, 2, 3]
  - VMAF best; problem: reference usually not available
- for live/adaptive encoding:
  - fast, accurate, no-reference quality estimation

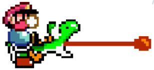
$\rightarrow$ nofu



Gaming QoE and video quality

- several influencing factors: *Möller, Schmidt, and Zadtootaghaj* [8]
  - video quality factors: content (cgi), encoding (fast),
  - interaction: delay, ...
- objective full-reference metrics: good results: *Barman et al.* [1, 2, 3]
  - VMAF best; problem: reference usually not available
- for live/adaptive encoding:
  - fast, accurate, no-reference quality estimation

→ nofu

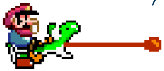


Gaming QoE and video quality

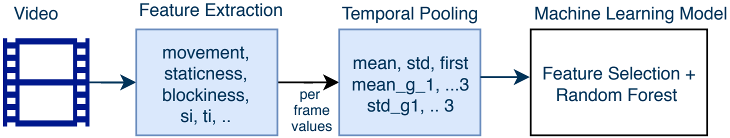
TECHNISCHE UNIVERSITÄT ILMENAU

- several influencing factors: *Möller, Schmidt, and Zadtootaghaj* [8]
  - video quality factors: content (cgi), encoding (fast),
  - interaction: delay, . . .
- objective full-reference metrics: good results: *Barman et al.* [1, 2, 3]
  - VMAF best; problem: reference usually not available
- for live/adaptive encoding:
  - fast, accurate, no-reference quality estimation

$\rightarrow$ nofu



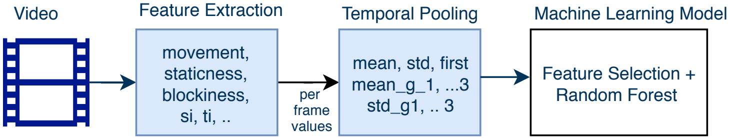
nofu – Features and Approach



- features:
  - $si^I + ti^M$ [6], fft^I [7], $staticness^I$, $blockiness^I$ [5],
  - cubrow-$\{first,last\}^M$, cubcol-$\{first,last\}^M$, $blockmotion^M$ [5]
- speedup: 360p center crop of input video
- temporal pooling: 12 feature values per frame
  - *first*, *mean*, *std*, groups $g = [1, 2, 3]$: $mean_g$, std_g
  - $\rightarrow$ duration independent 108 values per sequence
- ML algorithm: feature selection + RF
- additional no-ref model: brisque+niqe features, similar pipeline

$\rightarrow$ Evaluation and used Dataset

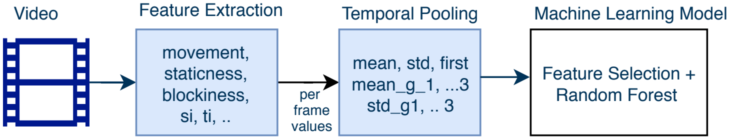
nofu – Features and Approach



- features:
  - $si^{I}+ti^{M}$ [6], fft^{I} [7], $staticness^{I}$, $blockiness^{I}$[5],
  - cubrow-{first,last}^{M}, cubcol-{first,last}^{M}, blockmotion^{M}[5]
- speedup: 360p center crop of input video
- temporal pooling: 12 feature values per frame
  - *first*, *mean*, *std*, groups $g = [1, 2, 3]$: $mean_g$, std_g
  - → duration independent 108 values per sequence
- ML algorithm: feature selection + RF
- additional no-ref model: brisque+niqe features, similar pipeline

→ Evaluation and used Dataset

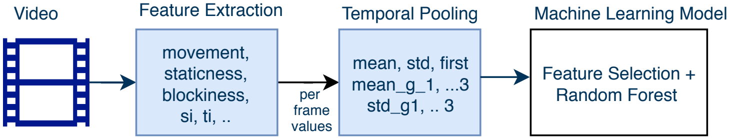
nofu – Features and Approach



- features:
  - si^I+ti^M [6], fft^I [7], staticness^I, blockiness^I[5],
  - cubrow-{first,last}^{M}, cubcol-{first,last}^{M}, blockmotion^M[5]
- speedup: 360p center crop of input video
- temporal pooling: 12 feature values per frame
  - *first*, *mean*, *std*, groups $g = [1, 2, 3]$: $mean_g$, std_g
  - $\rightarrow$ duration independent 108 values per sequence
- ML algorithm: feature selection + RF
- additional no-ref model: brisque+niqe features, similar pipeline

$\rightarrow$ Evaluation and used Dataset

nofu – Features and Approach

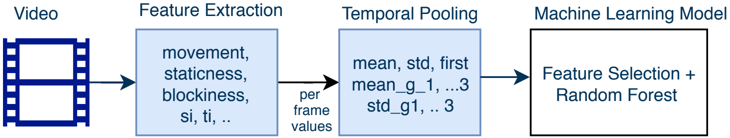


- features:
  - $\text{si}^{I}+\text{ti}^{M}$ [6], fft^{I} [7], staticness^{I}, blockiness^{I}[5],
  - cubrow-$\{$first,last$\}^{M}$, cubcol-$\{$first,last$\}^{M}$, blockmotion^{M}[5]
- speedup: 360p center crop of input video
- temporal pooling: 12 feature values per frame
  - *first*, *mean*, *std*, groups $g = [1, 2, 3]$: $mean_g$, std_g
  - $\rightarrow$ duration independent 108 values per sequence
- ML algorithm: feature selection + RF
- additional no-ref model: brisque+niqe features, similar pipeline

$\rightarrow$ Evaluation and used Dataset

nofu – Features and Approach

TECHNISCHE UNIVERSITÄT ILMENAU

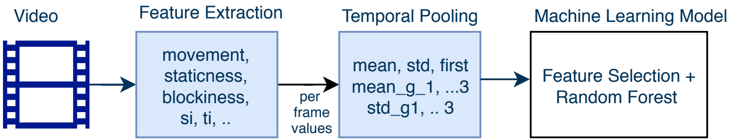


- features:
  - $si^{I}+ti^{M}$ [6], fft^{I} [7], $staticness^{I}$, $blockiness^{I}$[5],
  - cubrow-$\{first,last\}^{M}$, cubcol-$\{first,last\}^{M}$, $blockmotion^{M}$[5]
- speedup: 360p center crop of input video
- temporal pooling: 12 feature values per frame
  - *first*, *mean*, *std*, groups $g = [1, 2, 3]$: $mean_g$, std_g
  - $\rightarrow$ duration independent 108 values per sequence
- ML algorithm: feature selection + RF
- additional no-ref model: brisque+niqe features, similar pipeline

$\rightarrow$ Evaluation and used Dataset

nofu – Features and Approach

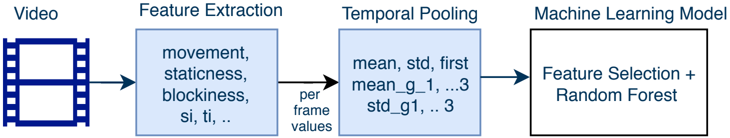
TECHNISCHE UNIVERSITÄT
ILMENAU



- features:
  - $\mathrm{si}^I + \mathrm{ti}^M$ [6], fft^I [7], $\mathrm{staticness}^I$, $\mathrm{blockiness}^I$[5],
  - cubrow-$\{$first,last$\}^M$, cubcol-$\{$first,last$\}^M$, $\mathrm{blockmotion}^M$[5]
- speedup: 360p center crop of input video
- temporal pooling: 12 feature values per frame
  - *first*, *mean*, *std*, groups $g = [1, 2, 3]$: $mean_g$, std_g
  - $\rightarrow$ duration independent 108 values per sequence
- ML algorithm: feature selection + RF
- additional no-ref model: brisque+niqe features, similar pipeline

$\rightarrow$ Evaluation and used Dataset

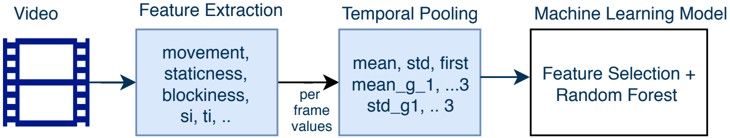
nofu – Features and Approach



- features:
  - si^I+ti^M [6], fft^I [7], staticness^I, blockiness^I[5],
  - cubrow-$\{$first,last$\}^M$, cubcol-$\{$first,last$\}^M$, blockmotion^M[5]
- speedup: 360p center crop of input video
- temporal pooling: 12 feature values per frame
  - *first*, *mean*, *std*, groups $g = [1, 2, 3]$: $mean_g$, std_g
  - $\rightarrow$ duration independent 108 values per sequence
- ML algorithm: feature selection + RF
- additional no-ref model: brisque+niqe features, similar pipeline

$\rightarrow$ Evaluation and used Dataset

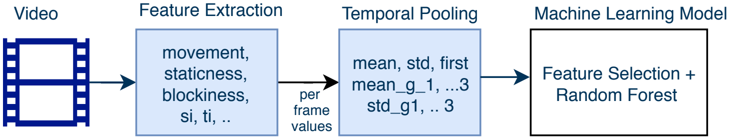
nofu – Features and Approach



- features:
  - si^I+ti^M [6], fft^I [7], staticness^I, blockiness^I[5],
  - cubrow-$\{$first,last$\}^M$, cubcol-$\{$first,last$\}^M$, blockmotion^M[5]
- speedup: 360p center crop of input video
- temporal pooling: 12 feature values per frame
  - *first*, *mean*, *std*, groups $g = [1, 2, 3]$: $mean_g$, std_g
  - $\rightarrow$ duration independent 108 values per sequence
- ML algorithm: feature selection + RF
- additional no-ref model: brisque+niqe features, similar pipeline

$\rightarrow$ Evaluation and used Dataset

nofu – Features and Approach

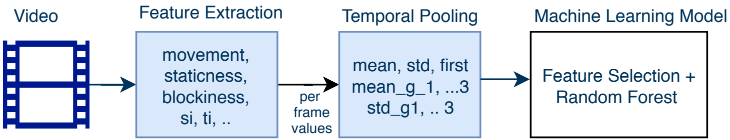


- features:
  - si^{I}+ti^{M} [6], fft^{I} [7], staticness^{I}, blockiness^{I}[5],
  - cubrow-$\{$first,last$\}^{M}$, cubcol-$\{$first,last$\}^{M}$, blockmotion^{M}[5]
- speedup: 360p center crop of input video
- temporal pooling: 12 feature values per frame
  - *first*, *mean*, *std*, groups $g = [1, 2, 3]$: $mean_g$, std_g
  - $\rightarrow$ duration independent 108 values per sequence
- ML algorithm: feature selection + RF
- additional no-ref model: brisque+niqe features, similar pipeline

$\rightarrow$ Evaluation and used Dataset

nofu – Features and Approach

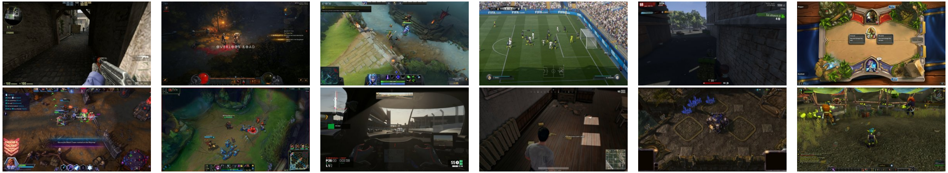
TECHNISCHE UNIVERSITÄT ILMENAU



- features:
  - si^I+ti^M [6], fft^I [7], staticness^I, blockiness^I[5],
  - cubrow-$\{$first,last$\}^M$, cubcol-$\{$first,last$\}^M$, blockmotion^M[5]
- speedup: 360p center crop of input video
- temporal pooling: 12 feature values per frame
  - *first*, *mean*, *std*, groups $g = [1, 2, 3]$: $mean_g$, std_g
  - $\rightarrow$ duration independent 108 values per sequence
- ML algorithm: feature selection + RF
- additional no-ref model: brisque+niqe features, similar pipeline

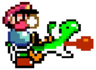
$\rightarrow$ Evaluation and used Dataset

Evaluation – Dataset

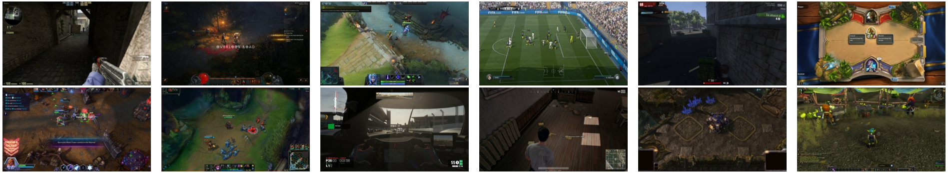


- GamingVideoSET: *Barman et al.* [4]:
  - 24 full-HD sources, 576 distorted videos, 90 with subjective scores
- two main evaluations: 10-fold cross validation and source fold:
  - (1) based on VMAF, (2) based on subjective scores

  → MOS prediction

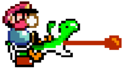


Evaluation – Dataset



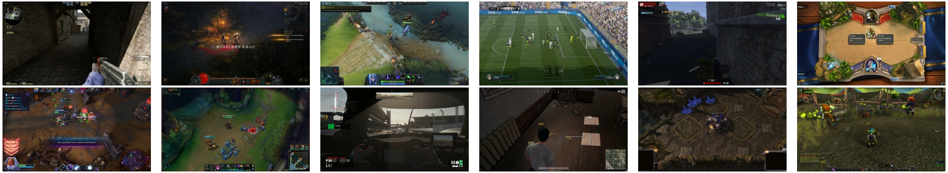
- GamingVideoSET: *Barman et al.* [4]:
  - 24 full-HD sources, 576 distorted videos, 90 with subjective scores
- two main evaluations: 10-fold cross validation and source fold:
  - (1) based on VMAF, (2) based on subjective scores

  → MOS prediction



Evaluation – Dataset

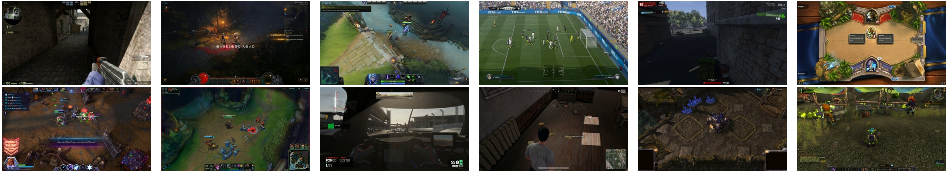
TECHNISCHE UNIVERSITÄT
ILMENAU



- GamingVideoSET: *Barman et al.* [4]:
  - 24 full-HD sources, 576 distorted videos, 90 with subjective scores
- two main evaluations: 10-fold cross validation and source fold:
  - (1) based on VMAF, (2) based on subjective scores

  → MOS prediction

Evaluation – Dataset

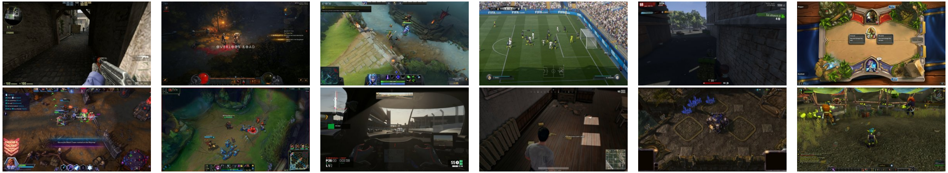


- GamingVideoSET: *Barman et al.* [4]:
  - 24 full-HD sources, 576 distorted videos, 90 with subjective scores
- two main evaluations: 10-fold cross validation and source fold:
  - (1) based on VMAF, (2) based on subjective scores

  → MOS prediction

Evaluation – Dataset

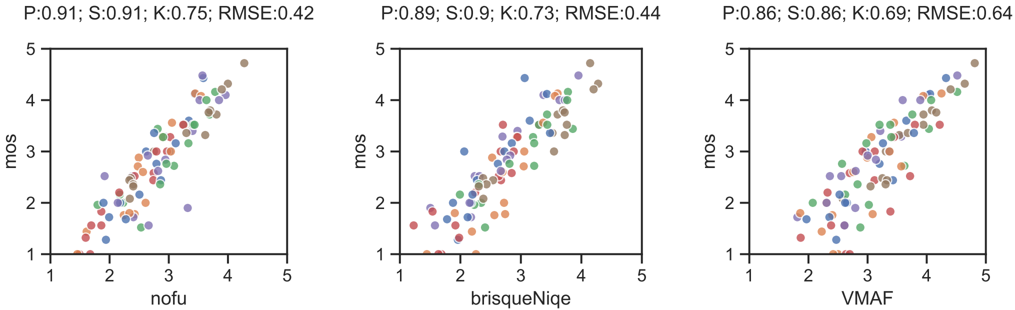
TECHNISCHE UNIVERSITÄT
ILMENAU



- GamingVideoSET: *Barman et al.* [4]:
  - 24 full-HD sources, 576 distorted videos, 90 with subjective scores
- two main evaluations: 10-fold cross validation and source fold:
  - (1) based on VMAF, (2) based on subjective scores

  $\rightarrow$ MOS prediction

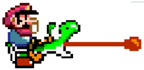
Evaluation – MOS prediction



pearson (P), spearman (S), kendall (K) and RMSE

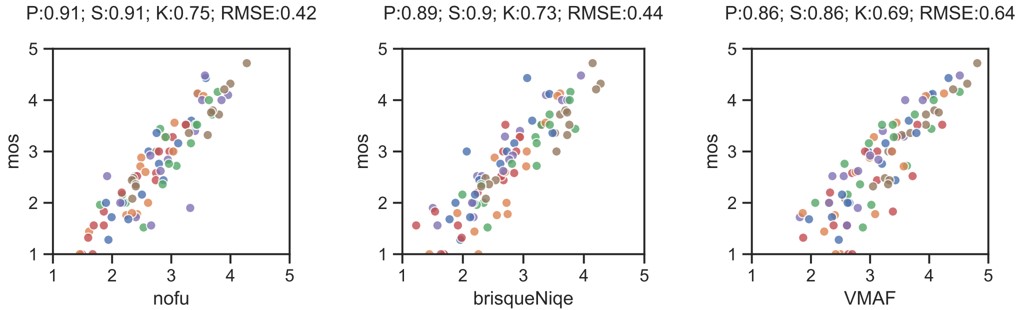
- nofu > brisque+niqe > vmaf > ssim
- source video fold evaluation: nofu > brisque+niqe

  → Conclusion



Evaluation – MOS prediction

TECHNISCHE UNIVERSITÄT
ILMENAU

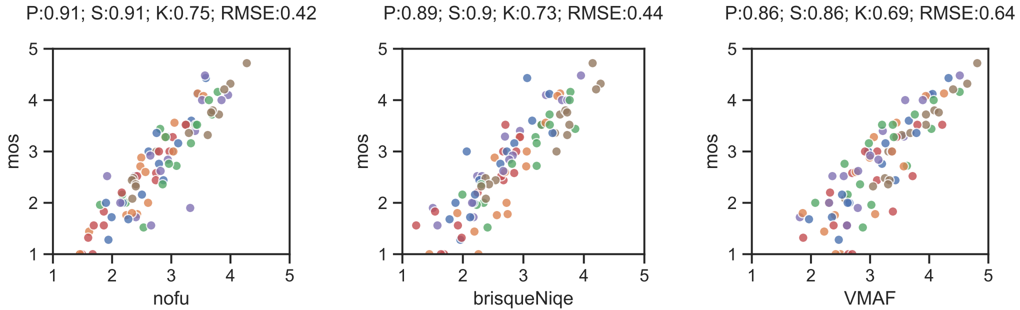


pearson (P), spearman (S), kendall (K) and RMSE

- nofu > brisque+niqe > vmaf > ssim
- source video fold evaluation: nofu > brisque+niqe

→ Conclusion

Evaluation – MOS prediction



pearson (P), spearman (S), kendall (K) and RMSE

- nofu > brisque+niqe > vmaf > ssim
- source video fold evaluation: nofu > brisque+niqe

  → Conclusion

Conclusion, Summary and Future Work

TECHNISCHE UNIVERSITÄT ILMENAU

- introduced **nofu**: no-reference model for gaming video quality
  - **features**: quality-related and gaming-specific
  - **temporal pooling + 360p center crop**
  - machine learning based
- evaluation using GamingVideoSET [4]
  - **nofu** outperforms other no-ref models + VMAF
  - per source fold: promising results
- open and next steps:
  - include delay/latency, bitstream features, combine **nofu**+brisque+niqe
  - use features/approach for different tasks

Conclusion, Summary and Future Work

- introduced **nofu**: no-reference model for gaming video quality
  - **features**: quality-related and gaming-specific
  - **temporal pooling + 360p center crop**
  - machine learning based
- evaluation using GamingVideoSET [4]
  - **nofu** outperforms other no-ref models + VMAF
  - per source fold: promising results
- open and next steps:
  - include delay/latency, bitstream features, combine **nofu**+brisque+niqe
  - use features/approach for different tasks

Conclusion, Summary and Future Work

- introduced **nofu**: no-reference model for gaming video quality
  - **features**: quality-related and gaming-specific
  - **temporal pooling + 360p center crop**
  - machine learning based
- evaluation using GamingVideoSET [4]
  - **nofu** outperforms other no-ref models + VMAF
  - per source fold: promising results
- open and next steps:
  - include delay/latency, bitstream features, combine **nofu**+brisque+niqe
  - use features/approach for different tasks

Conclusion, Summary and Future Work

TECHNISCHE UNIVERSITÄT **ILMENAU**

- introduced **nofu**: no-reference model for gaming video quality
  - **features**: quality-related and gaming-specific
  - **temporal pooling + 360p center crop**
  - machine learning based
- evaluation using GamingVideoSET [4]
  - **nofu** outperforms other no-ref models + VMAF
  - per source fold: promising results
- open and next steps:
  - include delay/latency, bitstream features, combine **nofu**+brisque+niqe
  - use features/approach for different tasks

Conclusion, Summary and Future Work

TECHNISCHE UNIVERSITÄT
ILMENAU

- introduced **nofu**: no-reference model for gaming video quality
  - **features**: quality-related and gaming-specific
  - **temporal pooling + 360p center crop**
  - machine learning based
- evaluation using GamingVideoSET [4]
  - **nofu** outperforms other no-ref models + VMAF
  - per source fold: promising results
- open and next steps:
  - include delay/latency, bitstream features, combine **nofu**+brisque+niqe
  - use features/approach for different tasks

Conclusion, Summary and Future Work

- introduced **nofu**: no-reference model for gaming video quality
  - **features**: quality-related and gaming-specific
  - **temporal pooling + 360p center crop**
  - machine learning based
- evaluation using GamingVideoSET [4]
  - **nofu** outperforms other no-ref models + VMAF
  - per source fold: promising results
- open and next steps:
  - include delay/latency, bitstream features, combine **nofu**+brisque+niqe
  - use features/approach for different tasks

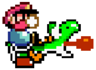
Conclusion, Summary and Future Work

TECHNISCHE UNIVERSITÄT
ILMENAU

- introduced **nofu**: no-reference model for gaming video quality
  - **features**: quality-related and gaming-specific
  - **temporal pooling + 360p center crop**
  - machine learning based
- evaluation using GamingVideoSET [4]
  - **nofu** outperforms other no-ref models + VMAF
  - per source fold: promising results
- open and next steps:
  - include delay/latency, bitstream features, combine **nofu**+brisque+niqe
  - use features/approach for different tasks

Conclusion, Summary and Future Work

- introduced **nofu**: no-reference model for gaming video quality
  - **features**: quality-related and gaming-specific
  - **temporal pooling + 360p center crop**
  - machine learning based
- evaluation using GamingVideoSET [4]
  - **nofu** outperforms other no-ref models + VMAF
  - per source fold: promising results
- open and next steps:
  - include delay/latency, bitstream features, combine **nofu**+brisque+niqe
  - use features/approach for different tasks



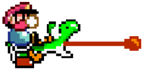
Conclusion, Summary and Future Work

TECHNISCHE UNIVERSITÄT
ILMENAU

- introduced **nofu**: no-reference model for gaming video quality
  - **features**: quality-related and gaming-specific
  - **temporal pooling + 360p center crop**
  - machine learning based
- evaluation using GamingVideoSET [4]
  - **nofu** outperforms other no-ref models + VMAF
  - per source fold: promising results
- open and next steps:
  - include delay/latency, bitstream features, combine **nofu**+brisque+niqe
  - use features/approach for different tasks

Conclusion, Summary and Future Work

- introduced **nofu**: no-reference model for gaming video quality
  - **features**: quality-related and gaming-specific
  - **temporal pooling + 360p center crop**
  - machine learning based
- evaluation using GamingVideoSET [4]
  - **nofu** outperforms other no-ref models + VMAF
  - per source fold: promising results
- open and next steps:
  - include delay/latency, bitstream features, combine **nofu**+brisque+niqe
  - use features/approach for different tasks



Thank you for your attention

...... are there any questions?

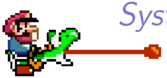
References I

[1] Nabajeet Barman and Maria G Martini. “H. 264/MPEG-AVC, H. 265/MPEG-HEVC and VP9 codec comparison for live gaming video streaming”. In: *Quality of Multimedia Experience (QoMEX), 2017 Ninth International Conference on*. IEEE. IEEE, 2017, pp. 1–6.

[2] Nabajeet Barman et al. “A Comparative Quality Assessment Study for Gaming and Non-Gaming Videos”. In: *2018 Tenth International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE. 2018, pp. 1–6.

[3] Nabajeet Barman et al. “An evaluation of video quality assessment metrics for passive gaming video streaming”. In: *Proceedings of the 23rd Packet Video Workshop*. ACM. 2018, pp. 7–12.

[4] Nabajeet Barman et al. “GamingVideoSET: a dataset for gaming video streaming applications”. In: *2018 16th Annual Workshop on Network and Systems Support for Games (NetGames)*. IEEE. 2018, pp. 1–6.



References II

[5] Steve Göring et al. “Analyze And Predict the Perceptibility of UHD Video Contents”. In: *Electronic Imaging, Human Vision Electronic Imaging* (2019).

[6] ITU-T. *Subjective video quality assessment methods for multimedia applications*. Serie P: Telephone Transmission Quality, Telephone Installations, Local Line Networks. Vol. Recommendation ITU-T P.910. International Telecommunication Union. Geneva, 2008.

[7] Ioannis Katsavounidis, Anne Aaron, and David Ronca. “Native resolution detection of video sequences”. In: *Annual Technical Conference and Exhibition, SMPTE 2015*. SMPTE. 2015, pp. 1–20.

[8] Sebastodes Möller, Steven Schmidt, and Saman Zadtootaghaj. “New ITU-T Standards for Gaming QoE Evaluation and Management”. In: *2018 Tenth International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE. 2018, pp. 1–6.

References III

[9] Saman Zadtootaghaj et al. “A classification of video games based on game characteristics linked to video coding complexity”. In: *2018 16th Annual Workshop on Network and Systems Support for Games (NetGames)*. IEEE. 2018, pp. 1–6.