DeViQ – A deep no reference video quality model

Steve Göring Janto Skowronek Alexander Raake

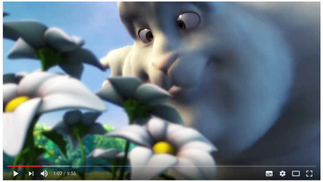
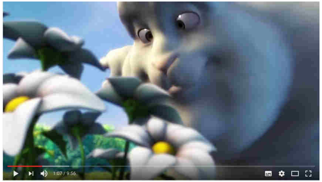
Audiovisual Technology Group, Technische Universität Ilmenau, Germany;
Email: [steve.goering, janto.skowronek, alexander.raake]@tu-ilmenau.de,

March 22, 2018

TECHNISCHE UNIVERSITÄT
ILMENAU

Motivation

TECHNISCHE UNIVERSITÄT
ILMENAU

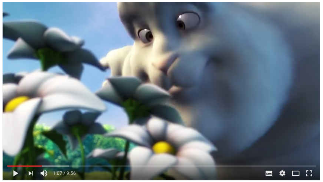
  

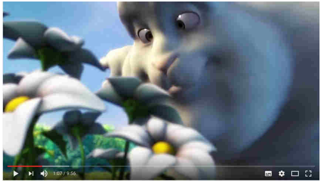
- most internet traffic generated via video streaming providers [4]
- user's expectation: best possible video quality under every condition
- trending technologies: **4k/UHD**, HDR, 360 degree, encoders, ...
- automated monitoring/optimization of perceived video quality

→ a brief look on current pixel based video/image quality models

Motivation

TECHNISCHE UNIVERSITÄT
ILMENAU



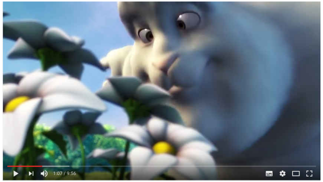


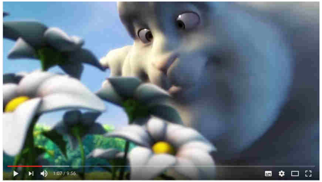
- most internet traffic generated via video streaming providers [4]
- user's expectation: best possible video quality under every condition
- trending technologies: **4k/UHD**, HDR, 360 degree, encoders, ...
- automated monitoring/optimization of perceived video quality

→ a brief look on current pixel based video/image quality models

Motivation

TECHNISCHE UNIVERSITÄT ILMENAU



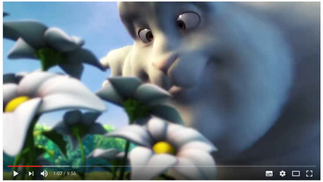


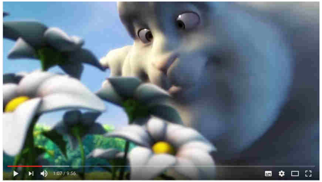
- most internet traffic generated via video streaming providers [4]
- user's expectation: best possible video quality under every condition
- trending technologies: **4k/UHD**, HDR, 360 degree, encoders, ...
- automated monitoring/optimization of perceived video quality

→ a brief look on current pixel based video/image quality models

Motivation

TECHNISCHE UNIVERSITÄT
ILMENAU



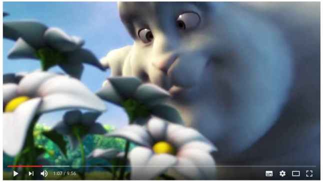


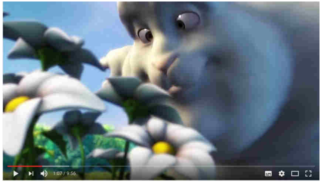
- most internet traffic generated via video streaming providers [4]
- user's expectation: best possible video quality under every condition
- trending technologies: **4k/UHD**, HDR, 360 degree, encoders, ...
- automated monitoring/optimization of perceived video quality

→ a brief look on current pixel based video/image quality models

Motivation

TECHNISCHE UNIVERSITÄT ILMENAU





- most internet traffic generated via video streaming providers [4]
- user's expectation: best possible video quality under every condition
- trending technologies: **4k/UHD**, HDR, 360 degree, encoders, ...
- automated monitoring/optimization of perceived video quality

→ a brief look on current pixel based video/image quality models

Video/image quality models

TECHNISCHE UNIVERSITÄT ILMENAU

- full-reference models highly accurate to human perception [18]
  - e.g. Netflix's VMAF [14] → **reference video**
- hand-crafted features [12, 14]
  - new encoders/ technologies → new artefacts → **new features**
- models using deep neural networks [3, 11, 8, 5, 6, 9]
  - patching to reduce input size → **losses global connections; many patches for 4K**
  - training requires per frame quality scores → **huge database**

→ ideas for solving the identified research problems

Video/image quality models

TECHNISCHE UNIVERSITÄT ILMENAU

- full-reference models highly accurate to human perception [18]
  - e.g. Netflix's VMAF [14] → **reference video**
- hand-crafted features [12, 14]
  - new encoders/ technologies → new artefacts → **new features**
- models using deep neural networks [3, 11, 8, 5, 6, 9]
  - patching to reduce input size → **losses global connections; many patches for 4K**
  - training requires per frame quality scores → **huge database**

→ ideas for solving the identified research problems

Video/image quality models

TECHNISCHE UNIVERSITÄT ILMENAU

- full-reference models highly accurate to human perception [18]
  - e.g. Netflix's VMAF [14] → **reference video**
- hand-crafted features [12, 14]
  - new encoders/ technologies → new artefacts → **new features**
- models using deep neural networks [3, 11, 8, 5, 6, 9]
  - patching to reduce input size → **losses global connections; many patches for 4K**
  - training requires per frame quality scores → **huge database**

→ ideas for solving the identified research problems

Video/image quality models

- full-reference models highly accurate to human perception [18]
  - e.g. Netflix's VMAF [14] → **reference video**
- hand-crafted features [12, 14]
  - new encoders/ technologies → new artefacts → **new features**
- models using deep neural networks [3, 11, 8, 5, 6, 9]
  - patching to reduce input size → **losses global connections; many patches for 4K**
  - training requires per frame quality scores → **huge database**

→ ideas for solving the identified research problems

Video/image quality models

- full-reference models highly accurate to human perception [18]
  - e.g. Netflix's VMAF [14] → **reference video**
- hand-crafted features [12, 14]
  - new encoders/ technologies → new artefacts → **new features**
- models using deep neural networks [3, 11, 8, 5, 6, 9]
  - patching to reduce input size → **losses global connections; many patches for 4K**
  - training requires per frame quality scores → **huge database**

→ ideas for solving the identified research problems

Video/image quality models

TECHNISCHE UNIVERSITÄT
ILMENAU

- full-reference models highly accurate to human perception [18]
  - e.g. Netflix's VMAF [14] → **reference video**
- hand-crafted features [12, 14]
  - new encoders/ technologies → new artefacts → **new features**
- models using deep neural networks [3, 11, 8, 5, 6, 9]
  - patching to reduce input size → **losses global connections; many patches for 4K**
  - training requires per frame quality scores → **huge database**

→ ideas for solving the identified research problems

Video/image quality models

TECHNISCHE UNIVERSITÄT ILMENAU

- full-reference models highly accurate to human perception [18]
  - e.g. Netflix's VMAF [14] → **reference video**
- hand-crafted features [12, 14]
  - new encoders/ technologies → new artefacts → **new features**
- models using deep neural networks [3, 11, 8, 5, 6, 9]
  - patching to reduce input size → **losses global connections; many patches for 4K**
  - training requires per frame quality scores → **huge database**

→ ideas for solving the identified research problems

Video/image quality models

TECHNISCHE UNIVERSITÄT
ILMENAU

- full-reference models highly accurate to human perception [18]
  - e.g. Netflix's VMAF [14] → **reference video**
- hand-crafted features [12, 14]
  - new encoders/ technologies → new artefacts → **new features**
- models using deep neural networks [3, 11, 8, 5, 6, 9]
  - patching to reduce input size → **losses global connections; many patches for 4K**
  - training requires per frame quality scores → **huge database**

→ ideas for solving the identified research problems

Video/image quality models

TECHNISCHE UNIVERSITÄT
ILMENAU

- full-reference models highly accurate to human perception [18]
  - e.g. Netflix's VMAF [14] → **reference video**
- hand-crafted features [12, 14]
  - new encoders/ technologies → new artefacts → **new features**
- models using deep neural networks [3, 11, 8, 5, 6, 9]
  - patching to reduce input size → **losses global connections; many patches for 4K**
  - training requires per frame quality scores → **huge database**

→ ideas for solving the identified research problems

How to solve the identified problems?

- huge training database for no-reference model:
  - generate ground-truth per frame data from full-reference model: VMAF [14, 10]
- hand-crafted features
  - using a pre-trained DNN for automatic feature extraction: inception-v3 [17]
- patching and global connection; many patches for 4K resolution
  - using hierarchical sub-images with larger block size: 299x299

→ introduce our model DeViQ (**De**ep **Vi**deo **Q**uality)

How to solve the identified problems?

TECHNISCHE UNIVERSITÄT ILMENAU

- huge training database for no-reference model:
  - generate ground-truth per frame data from full-reference model: VMAF [14, 10]
- hand-crafted features
  - using a pre-trained DNN for automatic feature extraction: inception-v3 [17]
- patching and global connection; many patches for 4K resolution
  - using hierarchical sub-images with larger block size: 299x299

→ introduce our model DeViQ (**De**ep **Vi**deo **Q**uality)

How to solve the identified problems?

TECHNISCHE UNIVERSITÄT
ILMENAU

- huge training database for no-reference model:
  - generate ground-truth per frame data from full-reference model: VMAF [14, 10]
- hand-crafted features
  - using a pre-trained DNN for automatic feature extraction: inception-v3 [17]
- patching and global connection; many patches for 4K resolution
  - using hierarchical sub-images with larger block size: 299x299

→ introduce our model DeViQ (**De**ep **Vi**deo **Q**uality)

How to solve the identified problems?

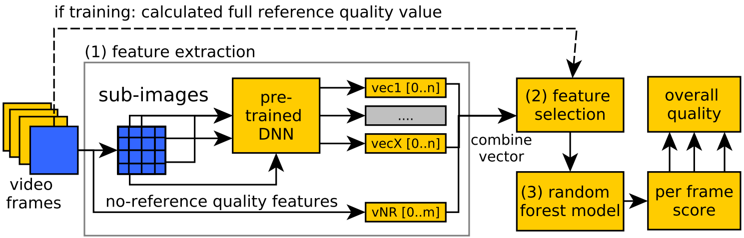
TECHNISCHE UNIVERSITÄT
ILMENAU

- huge training database for no-reference model:
  - generate ground-truth per frame data from full-reference model: VMAF [14, 10]
- hand-crafted features
  - using a pre-trained DNN for automatic feature extraction: inception-v3 [17]
- patching and global connection; many patches for 4K resolution
  - using hierarchical sub-images with larger block size: 299x299

→ introduce our model DeViQ (**De**ep **Vi**deo **Q**uality)

DeViQ– General approach

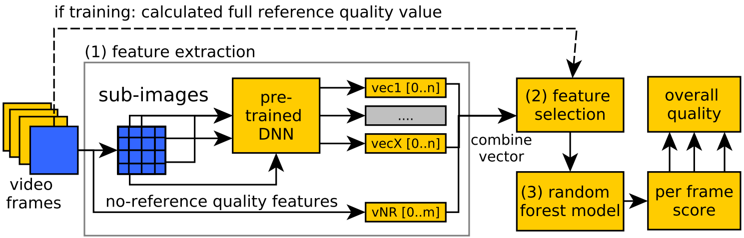
TECHNISCHE UNIVERSITÄT
ILMENAU



- (1) automatic feature extraction
  - pre-trained classification DNN
  - hierarchical sub-images: full, 1/2 of each dimension, 1/4 and 1/8 = 85 images
  - no-reference features; brisque+niqe [12, 13]
- (3) random forest model with (2) feature selection
- final quality score: mean value of each frame

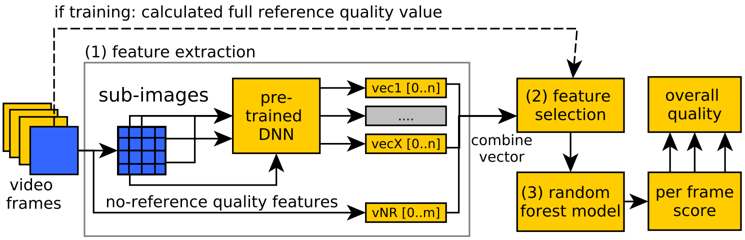
DeViQ– General approach

TECHNISCHE UNIVERSITÄT
ILMENAU



- (1) automatic feature extraction
  - pre-trained classification DNN
  - hierarchical sub-images: full, $1/2$ of each dimension, $1/4$ and $1/8 = 85$ images
  - no-reference features; brisque+niqe [12, 13]
- (3) random forest model with (2) feature selection
- final quality score: mean value of each frame

DeViQ– General approach



- (1) automatic feature extraction
  - pre-trained classification DNN
  - hierarchical sub-images: full, $1/2$ of each dimension, $1/4$ and $1/8 = 85$ images
  - no-reference features; brisque+niqe [12, 13]
- (3) random forest model with (2) feature selection
- final quality score: mean value of each frame

DeViQ – Evaluation – Dataset – Source Sequences

all source videos: UHD-I (3840x2160); 60 fps (except sintel*); 10 s

train


harmonic [7]


blender [2]*


TUIL


TUIL

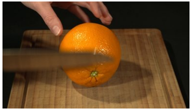

Sony [15]


Netflix

validation


harmonic [7]


blender [1]


TUIL


TUIL

Sony [16]


Netflix

DeViQ – Evaluation – Conditions

TECHNISCHE UNIVERSITÄT ILMENAU

- 3 codecs; 5 resolutions; 2/3 bitrates per resolution
- $\rightarrow$ encoded to 320 videos: **train=50%; validation=50%; no overlapp**
- calculated VMAF scores for $\approx 200k$ **frames**
- for validation: subjective test (22 participants; avg. age=26.7)
- comparison to retrained brisque+niqe model/ full-reference metrics

DeViQ – Evaluation – Conditions

TECHNISCHE UNIVERSITÄT
ILMENAU

- 3 codecs; 5 resolutions; 2/3 bitrates per resolution
- → encoded to 320 videos: **train=50%; validation=50%; no overlapp**
- calculated VMAF scores for $\approx 200k$ **frames**
- for validation: subjective test (22 participants; avg. age=26.7)
- comparison to retrained brisque+niqe model/ full-reference metrics

DeViQ – Evaluation – Conditions

TECHNISCHE UNIVERSITÄT ILMENAU

- 3 codecs; 5 resolutions; 2/3 bitrates per resolution
- $\rightarrow$ encoded to 320 videos: **train=50%; validation=50%; no overlapp**
- calculated VMAF scores for $\approx 200k$ **frames**
- for validation: subjective test (22 participants; avg. age=26.7)
- comparison to retrained brisque+niqe model/ full-reference metrics

DeViQ – Evaluation – Conditions

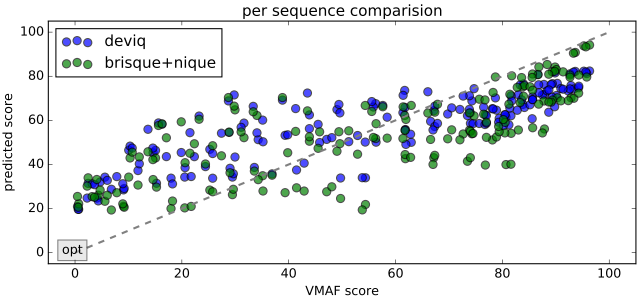
- 3 codecs; 5 resolutions; $2/3$ bitrates per resolution
- $\rightarrow$ encoded to 320 videos: **train=50%; validation=50%; no overlapp**
- calculated VMAF scores for $\approx 200k$ **frames**
- for validation: subjective test (22 participants; avg. age$=26.7$)
- comparison to retrained brisque+niqe model/ full-reference metrics

DeViQ – Evaluation – Conditions

- 3 codecs; 5 resolutions; $2/3$ bitrates per resolution
- $\rightarrow$ encoded to 320 videos: **train=50%; validation=50%; no overlapp**
- calculated VMAF scores for $\approx 200k$ **frames**
- for validation: subjective test (22 participants; avg. age$=26.7$)
- comparison to retrained brisque$+$niqe model/ full-reference metrics

DeViQ – Evaluation – Prediction vs. VMAF

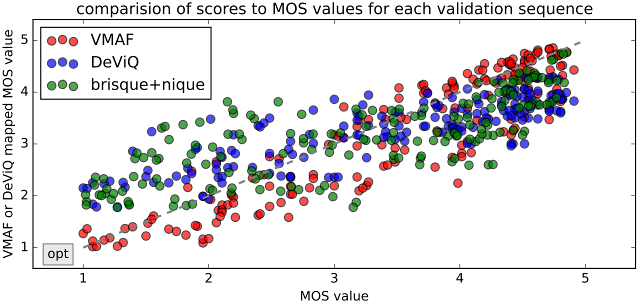
average VMAF-scores with DeViQ and brisque+niqe predictions



method	*RMSE*	R^2	pearson	kendall	spearman
deviq	18.87	0.60	0.84	0.66	0.84
brisque+nique	19.75	0.56	0.85	0.64	0.83
vifp	22.28	0.44	0.58	0.46	0.63

DeViQ – Evaluation – Prediction vs. MOS

comparison of VMAF, DeViQ, brisque+niqe to MOS values



method	$RMSE$	R^2	kendall	pearson	spearman
vmaf	0.55	0.76	0.72	0.92	0.89
deviq	0.70	0.61	0.61	0.84	0.81
brisque+nique	0.81	0.47	0.53	0.75	0.73
vifp	0.86	0.41	0.52	0.70	0.67

Conclusion and Future Work

TECHNISCHE UNIVERSITÄT ILMENAU

- identified main problems: hand-crafted features; patching; huge database $\rightarrow$ DeViQ (**De**ep **Vi**deo **Q**uality)
  - $\rightarrow$ performs good compared to full-reference, no-reference models
- open points:
  - frame and sub-image selection
  - average for overall video quality
- DeViQ's core idea: train a no-reference model based on a full-reference model using a pre-trained image DNN

Conclusion and Future Work

TECHNISCHE UNIVERSITÄT ILMENAU

- identified main problems: hand-crafted features; patching; huge database $\rightarrow$ DeViQ (**De**ep **Vi**deo **Q**uality)
  - $\rightarrow$ performs good compared to full-reference, no-reference models
- open points:
  - frame and sub-image selection
  - average for overall video quality
- DeViQ's core idea: train a no-reference model based on a full-reference model using a pre-trained image DNN

Conclusion and Future Work

TECHNISCHE UNIVERSITÄT
ILMENAU

- identified main problems: hand-crafted features; patching; huge database $\rightarrow$ DeViQ (**De**ep **Vi**deo **Q**uality)
  - $\rightarrow$ performs good compared to full-reference, no-reference models
- open points:
  - frame and sub-image selection
  - average for overall video quality
- DeViQ's core idea: train a no-reference model based on a full-reference model using a pre-trained image DNN

Conclusion and Future Work

TECHNISCHE UNIVERSITÄT
ILMENAU

- identified main problems: hand-crafted features; patching; huge database $\rightarrow$ DeViQ (**De**ep **Vi**deo **Q**uality)
  - $\rightarrow$ performs good compared to full-reference, no-reference models
- open points:
  - frame and sub-image selection
  - average for overall video quality
- DeViQ's core idea: train a no-reference model based on a full-reference model using a pre-trained image DNN

Conclusion and Future Work

- identified main problems: hand-crafted features; patching; huge database $\rightarrow$ DeViQ (**De**ep **Vi**deo **Q**uality)
  - $\rightarrow$ performs good compared to full-reference, no-reference models
- open points:
  - frame and sub-image selection
  - average for overall video quality
- DeViQ's core idea: train a no-reference model based on a full-reference model using a pre-trained image DNN

Conclusion and Future Work

TECHNISCHE UNIVERSITÄT ILMENAU

- identified main problems: hand-crafted features; patching; huge database $\rightarrow$ DeViQ (**De**ep **Vi**deo **Q**uality)
  - $\rightarrow$ performs good compared to full-reference, no-reference models
- open points:
  - frame and sub-image selection
  - average for overall video quality
- DeViQ's core idea: train a no-reference model based on a full-reference model using a pre-trained image DNN

Thank you for your attention

TECHNISCHE UNIVERSITÄT ILMENAU

...... are there any questions?

References I

[1] Blender Foundation. *Bick Buck Bunny Distribution*. URL: http://distribution.bbb3d.renderfarming.net/video/png (visited on 07/07/2017).

[2] Blender Foundation. *Sintel, the Durian Open Movie Project*. URL: https://media.xiph.org/sintel/sintel-4k-tiff16/ (visited on 07/07/2017).

[3] Sebastian Bosse et al. “Neural network-based full-reference image quality assessment”. In: *Picture Coding Symposium (PCS), 2016*. IEEE. 2016, pp. 1–5.

[4] Cisco. *Whitepaper: Cisco Visual Networking Index:Forecast and Methodology, 2015-2020*. 2015.

References II

[5] Prajna Paramita Dash, Akshaya Mishra, and Alexander Wong. "Deep Quality: A Deep No-reference Quality Assessment System". In: *arXiv preprint arXiv:1609.07170* (2016).

[6] Prajna Paramita Dash, Alexander Wong, and Akshaya Mishra. "VeNICE: A very deep neural network approach to no-reference image assessment". In: *Industrial Technology (ICIT), 2017 IEEE International Conference on*. IEEE. 2017, pp. 1091–1096.

[7] Harmonic. *Free 4K Demo Footage - Ultra HD Demo Footage*. URL: https://www.harmonicinc.com/4k-demo-footage-download/ (visited on 07/07/2017).

[8] Le Kang et al. "Convolutional neural networks for no-reference image quality assessment". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2014, pp. 1733–1740.

References III

[9] Jie Li et al. "No-reference image quality assessment using Prewitt magnitude based on convolutional neural networks". In: *Signal, Image and Video Processing* 10.4 (2016), pp. 609–616.

[10] J. Y. Lin et al. "A fusion-based video quality assessment (fvqa) index". In: *Signal and Information Processing Association Annual Summit and Conference (APSIPA), 2014 Asia-Pacific*. Dec. 2014, pp. 1–5.

[11] Vladimir V Lukin et al. "Combining full-reference image visual quality metrics by neural network." In: *Human Vision and Electronic Imaging*. 2015, 93940K.

[12] Anish Mittal, Anush Krishna Moorthy, and Alan Conrad Bovik. "No-reference image quality assessment in the spatial domain". In: *IEEE Transactions on Image Processing* 21.12 (2012), pp. 4695–4708.

References IV

TECHNISCHE UNIVERSITÄT ILMENAU

[13] Anish Mittal, Rajiv Soundararajan, and Alan C Bovik. “Making a “completely blind” image quality analyzer”. In: *IEEE Signal Processing Letters* 20.3 (2013), pp. 209–212.

[14] Netflix. *Netflix VMAF*. URL: https://github.com/Netflix/vmaf (visited on 07/08/2017).

[15] Sony. *Camping in Nature*. URL: http://4kmedia.org/sony-camping-in-nature-4k-demo/ (visited on 07/07/2017).

[16] Sony. *Surfing*. URL: http://4kmedia.org/sony-surfing-uhd-4k-demo/ (visited on 07/07/2017).

[17] Christian Szegedy et al. “Rethinking the Inception Architecture for Computer Vision”. In: *CoRR* (2015).

References V

[18] Maria Torres Vega et al. “An experimental survey of no-reference video quality assessment methods”. In: *International Journal of Pervasive Computing and Communications* 12.1 (2016), pp. 66–86.

deviq,brisque+nique vs. vmaf

method	*RMSE*	R^2	pearson	kendall	spearman
deviq	18.87	0.60	0.84	0.66	0.84
brisque+nique	19.75	0.56	0.85	0.64	0.83
vifp	22.28	0.44	0.58	0.46	0.63
msssim	48.99	-1.70	0.54	0.46	0.63
ssim	49.88	-1.80	0.48	0.44	0.60
psnrhvs	56.09	-2.55	0.33	0.52	0.72

deviq,brisque+nique,vmaf vs. mos

method	$RMSE$	R^2	cohen_d	kendall	pearson	spearman
vmaf	0.55	0.76	0.24	0.72	0.92	0.89
deviq	0.70	0.61	0.19	0.61	0.84	0.81
brisque+nique	0.81	0.47	0.34	0.53	0.75	0.73
vifp	0.86	0.41	-0.34	0.52	0.70	0.67
msssim	1.70	-1.32	-1.72	0.46	0.69	0.61
ssim	1.74	-1.42	-1.76	0.45	0.65	0.60
psnrhvs	2.27	-3.15	0.30	0.60	0.34	0.76

feature importance

TECHNISCHE UNIVERSITÄT
ILMENAU

for each 1000 feature values we summed the feature importance of our model; subimage 85=no-reference features