Appeal prediction for AI up-scaled Images

Steve Göring, Rasmus Merten, Alexander Raake

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

Code&Data: `https://bit.ly/3Z8Xmf5`

December 11, 2024

TECHNISCHE UNIVERSITÄT
ILMENAU

Motivation

TECHNISCHE UNIVERSITÄT
ILMENAU

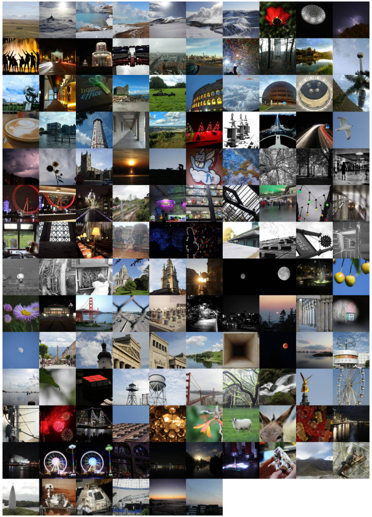
- AI-based image enhancement methods ...
  - e.g, for de-noising [14], in-painting [18], or re-colorization [20]
  - also for **up-scaling** [23, 3, 2, 24]
- typical comparison in state-of-the-art
  - objective models: PSNR/SSIM,
  - only one AI up-scaling method,
  - less often subjective evaluation
- our focus
  - which of the models is visually the best?
  - can the used method be detected after processing?
  - can the visual appeal predicted for up-scaling?

Upscaling Example

AI-based up-scaling examples; left: full image, then 270x270 pixels center crops of the source image, BSRGAN **x4** [24], and KXNet **x4** [3].

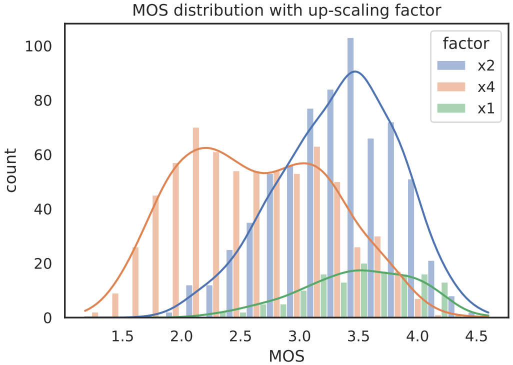
Dataset

- image up-scaling methods
  - BSRGAN [24], KXNet [3], Real-ESRGAN [23], waifu2x [16], **Lanczos**
  - two up-scaling factors (**x2**, **x4**)
- **high resolution (1080p) real content**
  - 136 source images
    - from "own" subset of **AVT-ImageAppeal-Dataset** [6]
    - re-scaling to 1080p height as reference **x1**;
    - for **x2** down-scaling to 540p; for **x4** to 270p
- total $136 \times \left(\underbrace{2}_{\textbf{x2,x4}} \times \underbrace{5}_{\text{methods}} + \underbrace{1}_{\textbf{x1}}\right) = 1496$



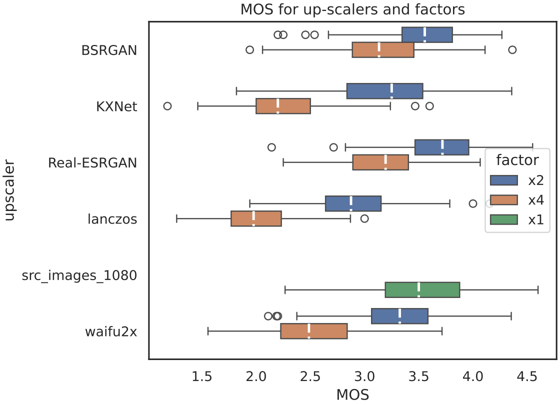
Evaluation - general

- online crowd-sourcing test
  - for image appeal
  - based on **AVRate Voyager** [10]
  - a participant rates 400 of 1496 images
- 55 participants (Clickworkers)
  - images have been rated
    - at least by 4,
    - at most by 25,
    - on average by 14.7 participants
- obvious: **x1** $>$ **x2** $>$ **x4**
- SOS analysis [12] $\rightarrow a \approx 0.275$



Evaluation - up-scaling algorithms

- best: Real-ESRGAN, BSRGAN
- worst: Lanczos, KXNet
- waifu2x in between



Up-scaled preferred over source image

Preference example; left: full source image; then 270x270 pixels center crops of the source image, Real-ESRGAN **x2**, and Real-ESRGAN **x4**.

- example source image: mean appeal rating of ≈ 3.78
- **x2** Real-ESRGAN variant: ≈ 4.0
- **x4** Real-ESRGAN: ≈ 3.17
- overall: upscaled $>$ source image: **x4**: 40%; **x2**: 74%

Detection

- which up-scaling method has been used?
  - multi-class classification, similar to [7, 5]
  - pre-trained baseline DNN,
  - transfer-learning [22], Keras [1]
  - images split into patches (224x224; no overlap)
  - 16 baseline models; 90%-10% train-validation
- best models:
  - DenseNet (f1-score $\approx$ 0.74) or ResNet variants
- worst: Inception or VGG variants
- Real-ESRGAN $\sim$ BSRGAN
- maybe better to predict: underlying generic method

Appeal prediction

- similar to detection
  - 16 DNNs, transfer-learning
  - regression instead of classification
  - mean appeal scores normalized to $[0, 1]$
  - center cropped inputs (224x224) [8, 4]
- DNN models
  - best: ResNet152V2, 0.83 PCC
  - worst: MobileNetV2, InceptionV3

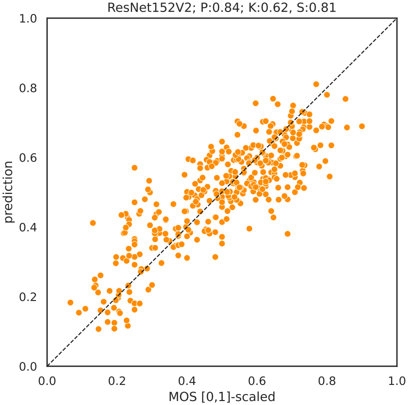


Image appeal compared to signal features

Feature	Pearson	Kendall	Spearman
combined	0.669	0.450	0.631
cpbd [17]	0.572	0.363	0.522
fft [11]	0.331	0.238	0.350
noise [11]	0.299	0.302	0.447
si	0.213	0.135	0.200
blur	0.152	0.110	0.164
saturation	0.093	0.057	0.087
colorfulness	0.024	0.019	0.028
contrast	-0.013	-0.005	-0.008
tone	-0.039	-0.020	-0.031
niqe	-0.088	-0.040	-0.061
blur strength	-0.378	-0.257	-0.375

implementation: [9]; ***combined***=RF model, 100 trees, scikit-learn [19]

Image appeal compared to SoA Models

TECHNISCHE UNIVERSITÄT
ILMENAU

Model	Pearson	Kendall	Spearman
DBCNN [25]	0.605	0.436	0.618
HYPERIQA [21]	0.601	0.414	0.592
CNNIQA	0.592	0.376	0.536
MUSIQ	0.555	0.365	0.522
MANIQA	0.505	0.345	0.493
paq2piq	0.492	0.311	0.448
NIMA quality CC [15]	0.433	0.281	0.408
ms_ssim	0.368	0.232	0.348
vif	0.363	0.248	0.373
psnr	0.248	0.164	0.244
ssim	0.183	0.135	0.203

implementation: PyTorch Image Quality (PIQ) Toolbox [13]

Conclusion, Summary and Future Work

TECHNISCHE UNIVERSITÄT
ILMENAU

- observation
  - DNN/AI-based up-scaling maybe better than traditional approaches
- open source dataset, subjective annotation, evaluation & comparison
  - 5 up-scaling methods, 2 factors, 1496 rated images
  - most appealing model Real-ESRGAN, second BSRGAN, Lanczos bad
  - reverse detection of which method used: possible
  - appeal prediction: new models needed, transfer learning promising
- future work
  - more tests with a larger number of source images/more methods
  - update/improve existing models to include AI distortions
  - video up-scaling with newer ai-based up-scaling methods

Thank you for your attention

TECHNISCHE UNIVERSITÄT
ILMENAU

...... are there any questions?

The authors would like to thank the participants for taking part in this crowd test. Furthermore, we want to thank the "**AG Wissenschaftliches Rechnen**" of the TU Ilmenau for providing computing resources.

References I

TECHNISCHE UNIVERSITÄT ILMENAU

[1] François Chollet et al. *Keras*. `https://keras.io`. 2015.

[2] Chao Dong et al. “Learning a deep convolutional network for image super-resolution”. In: *Computer Vision*. Springer. 2014, pp. 184–199.

[3] Jiahong Fu et al. “KXNet: A model-driven deep neural network for blind super-resolution”. In: *European Conference on Computer Vision*. Springer. 2022, pp. 235–253.

[4] Steve Göring, Christopher Krämmer, and Alexander Raake. “cencro – Speedup of Video Quality Calculation using Center Cropping”. In: *21st International Symposium on Multimedia (ISM)*. IEEE. Dec. 2019, pp. 1–8.

References II

[5] Steve Göring, Rasmus Merten, and Alexander Raake. "DNN-based Photography Rule Prediction using Photo Tags". In: *15th International Conference on Quality of Multimedia Experience (QoMEX)*. 2023.

[6] Steve Göring and Alexander Raake. "Image Appeal Revisited: Analysis, new Dataset and Prediction Models". In: *IEEE Access* 11 (2023), pp. 69563–69585.

[7] Steve Göring and Alexander Raake. "Rule of Thirds and Simplicity for Image Aesthetics using Deep Neural Networks". In: *23rd International Workshop on Multimedia Signal Processing (MMSP)*. IEEE. 2021, pp. 1–6.

References III

TECHNISCHE UNIVERSITÄT ILMENAU

[8] Steve Göring, Rakesh Rao Ramachandra Rao, and Alexander Raake. "Quality assessment of higher resolution images and videos with remote testing". In: *Quality and User Experience (QUEX)* 8.1 (2023), p. 2.

[9] Steve Göring et al. "Analysis of Appeal for realistic AI-generated Photos". In: *IEEE Access* 11 (2023), pp. 38999–39012.

[10] Steve Göring et al. "AVRate Voyager: an open source online testing platform". In: *23rd International Workshop on Multimedia Signal Processing (MMSP)*. IEEE. 2021, pp. 1–6.

[11] Steve Göring et al. "Modular Framework and Instances of Pixel-based Video Quality Models for UHD-1/4K". In: *IEEE Access* 9 (2021), pp. 31842–31864.

References IV

TECHNISCHE UNIVERSITÄT ILMENAU

[12] Tobias Hoßfeld, Raimund Schatz, and Sebastian Egger. "SOS: The MOS is not enough!" In: *3rd International Workshop on Quality of Multimedia Experience (QoMEX)*. IEEE. 2011, pp. 131–136.

[13] Sergey Kastryulin et al. *PyTorch Image Quality: Metrics for Image Quality Assessment*. 2022.

[14] Samuli Laine et al. "High-quality self-supervised deep image denoising". In: *Advances in Neural Information Processing Systems* 32 (2019).

[15] Christopher Lennan, Hao Nguyen, and Dat Tran. *Image Quality Assessment*. https://github.com/idealo/image-quality-assessment. 2018.

[16] nagadomi. *waifu2x – https://github.com/nagadomi/waifu2x*.

References V

[17] Niranjan D Narvekar and Lina J Karam. “A no-reference image blur metric based on the cumulative probability of blur detection (CPBD)”. In: *IEEE Transactions on Image Processing* 20.9 (2011), pp. 2678–2683.

[18] Deepak Pathak et al. “Context encoders: Feature learning by inpainting”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 2536–2544.

[19] F. Pedregosa et al. “Scikit-learn: Machine Learning in Python”. In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.

[20] Antoine Salmona, Lucía Bouza, and Julie Delon. “Deoldify: A review and implementation of an automatic colorization method”. In: *Image Processing On Line* 12 (2022), pp. 347–368.

References VI

TECHNISCHE UNIVERSITÄT ILMENAU

[21] Shaolin Su et al. "Blindly assess image quality in the wild guided by a self-adaptive hyper network". In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2020, pp. 3667–3676.

[22] Lisa Torrey and Jude Shavlik. "Transfer learning". In: *Handbook of research on machine learning applications and trends: algorithms, methods, and techniques*. IGI global, 2010, pp. 242–264.

[23] Xintao Wang et al. "Real-ESRGAN: Training Real-World Blind Super-Resolution with Pure Synthetic Data". In: *Proceedings of the IEEE/CVF international conference on computer vision*. 2021.

References VII

TECHNISCHE UNIVERSITÄT ILMENAU

[24] Kai Zhang et al. "Designing a practical degradation model for deep blind image super-resolution". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021, pp. 4791–4800.

[25] Weixia Zhang et al. "Blind Image Quality Assessment Using A Deep Bilinear Convolutional Neural Network". In: *Transaction on Circuits and Systems for Video Technology* 30.1 (2020), pp. 36–47.