A Loss Function for Generative Neural Networks Based on Watson's Perceptual Model

Steffen Czolbe[1] per.sc@di.ku.dk
Oswin Krause[1] oswin.krause@di.ku.dk
Ingemar Cox[1,2] ingemar.cox@du.ku.dk
Christian Igel[1] igel@du.ku.dk

[1]University of Copenhagen
[2]University College London

Summary

- Lossy compression algorithms (e.g. JPEG, MP3, MP4) discard components that fall below a perceptual threshold.
- We use a perceptual weighting inspired by these algorithms to introduce *Watson-DFT*, a reconstruction loss for VAEs.
- Compared to SSIM & MSE, VAEs trained with our metric generated sharper & more accurate images.
- We observed less artefacts, better generalization and fewer resource requirements compared to CNN-based loss functions.
- Representations learned on classification tasks may not translate well to generation tasks: Learned invariance to blurring and artifacts is not desired in image generation.

Watson's Perceptual Model

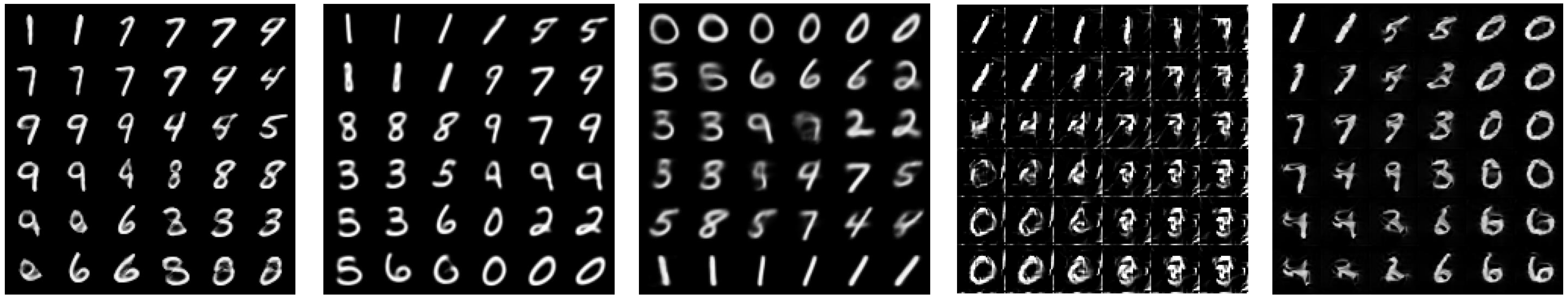
- Explicit model of the human visual system, used in image compression and watermarking [2].
- Light-weight, <150 parameters.
- Watson distance based on blockwise DCT coefficients C_{ijk} of K blocks sized $B \times B$, weighted by an image-dependent sensitivity table S:

$$D_{\text{Watson}}(C, C') = \sqrt[p]{\sum_{i,j,k=1}^{B,B,K} \left| \frac{C_{ijk} - C'_{ijk}}{S_{ijk}} \right|^p}$$

- We modified the model to be fully differentiable.
- To model spatial variations, we changed to the Discrete Fourier Transform (DFT).
- Let A be the DFT amplitudes and Φ the phase-information. We introduced weights $w_{ij} > 0$ and calculate the modified Watson's distance as:

$$L_{\text{Watson-DFT}}(A, \Phi, A', \Phi') = D_{\text{Watson}}(A, A') + \sum_{i,j,k=1}^{B,B,K} w_{ij} \arccos\left[\cos\left(\Phi_{ijk} - \Phi'_{ijk}\right)\right]$$

MNIST VAE Latent Manifolds

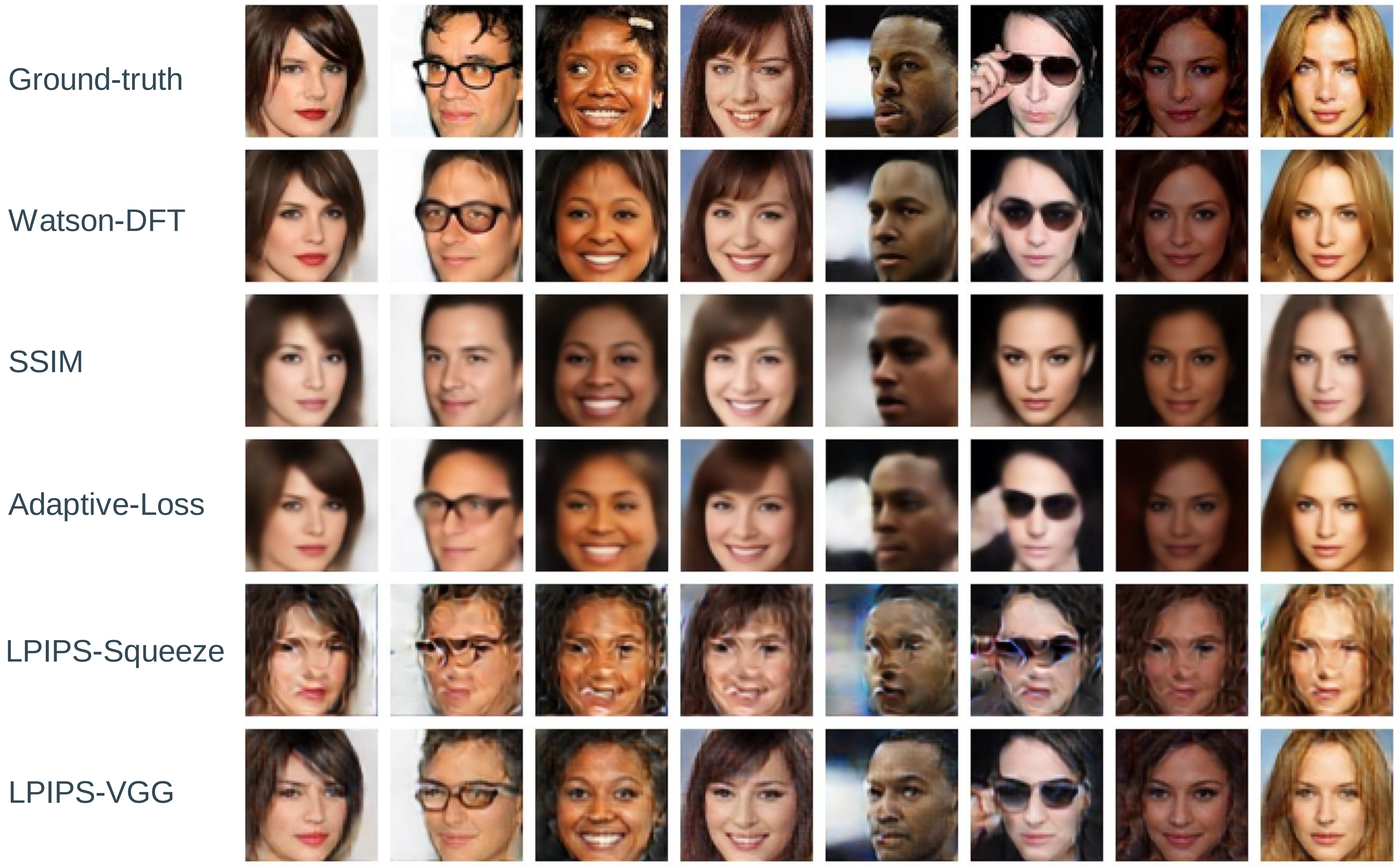


Watson-DFT SSIM Adaptive-Loss LPIPS-Squeeze LPIPS-VGG

Application to VAEs

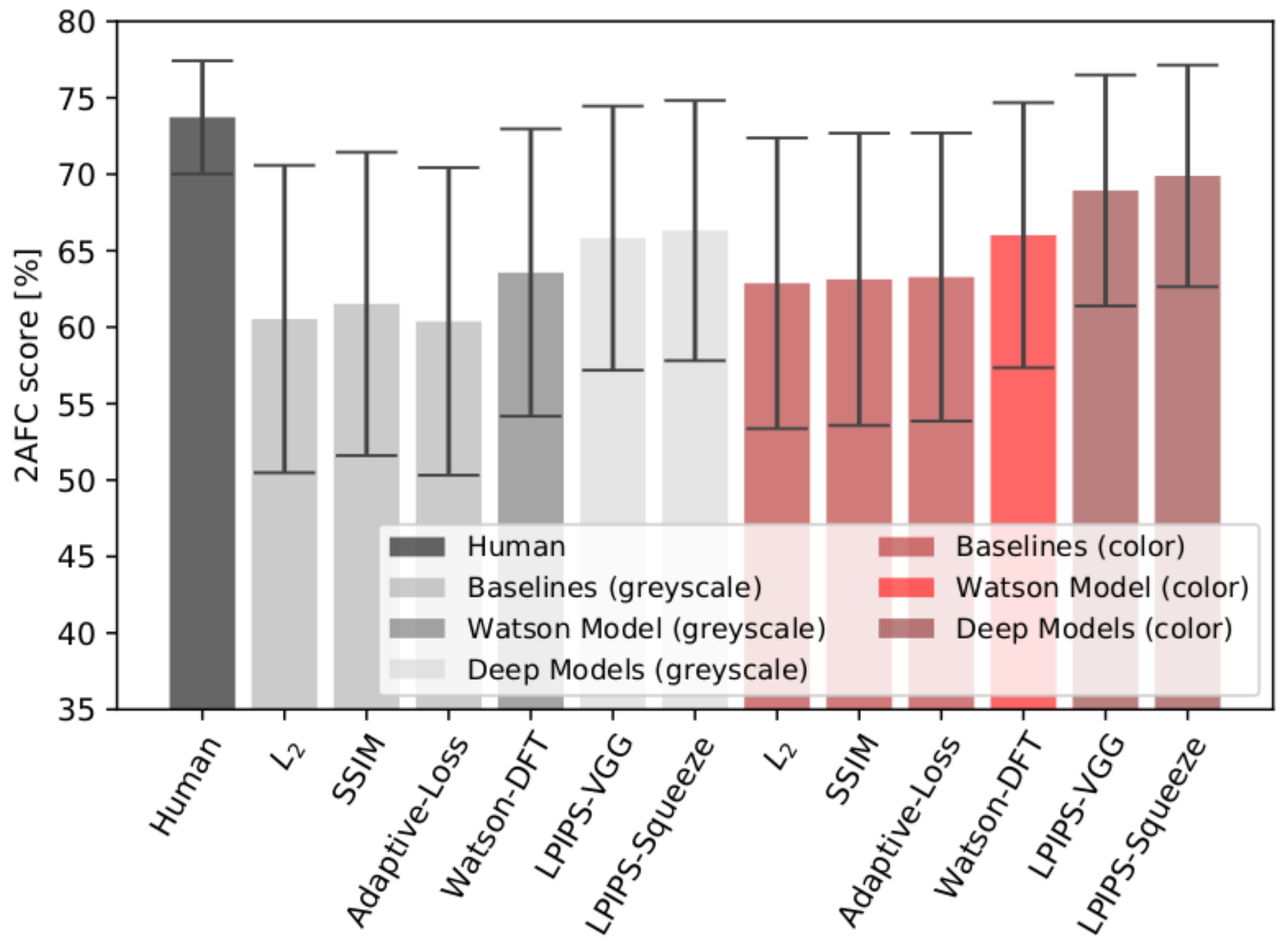
- We tuned the parameters of the loss function via gradient-based optimization on a dataset of human perceptual similarity measurements [3].
- We compared VAEs trained with our proposed Watson-DFT to SSIM, Adaptive-Loss [1] and LPIPS CNN-based loss functions [3].

CelebA VAE Reconstructions



Perceptual Score

- We evaluated agreement of the loss functions with the human similarity judgements [3].
- CNN-based loss functions achieved higher scores on this task, but did not lead to better VAEs for image generation.



References

[1] Jonathan T. Barron. "A general and adaptive robust loss function". In Conference on Computer Vision and Pattern Recognition (CVPR), 2019.

[2] Andrew B. Watson. "DCT quantization matrices visually optimized for individual images". In Human vision, visual processing, and digital display IV, vol. 1913, pp. 202–217. International Society for Optics and Photonics, 1993.

[3] Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. "The unreasonable effectiveness of deep features as a perceptual metric". In Conference on Computer Vision and Pattern Recognition (CVPR), 2018.