Perceptual impact of the loss function on deep-learning image coding performance

Shima Mohammadi, Joao Ascenso

Picture Coding Symposium, San Jose, California, USA, 7 December 2022

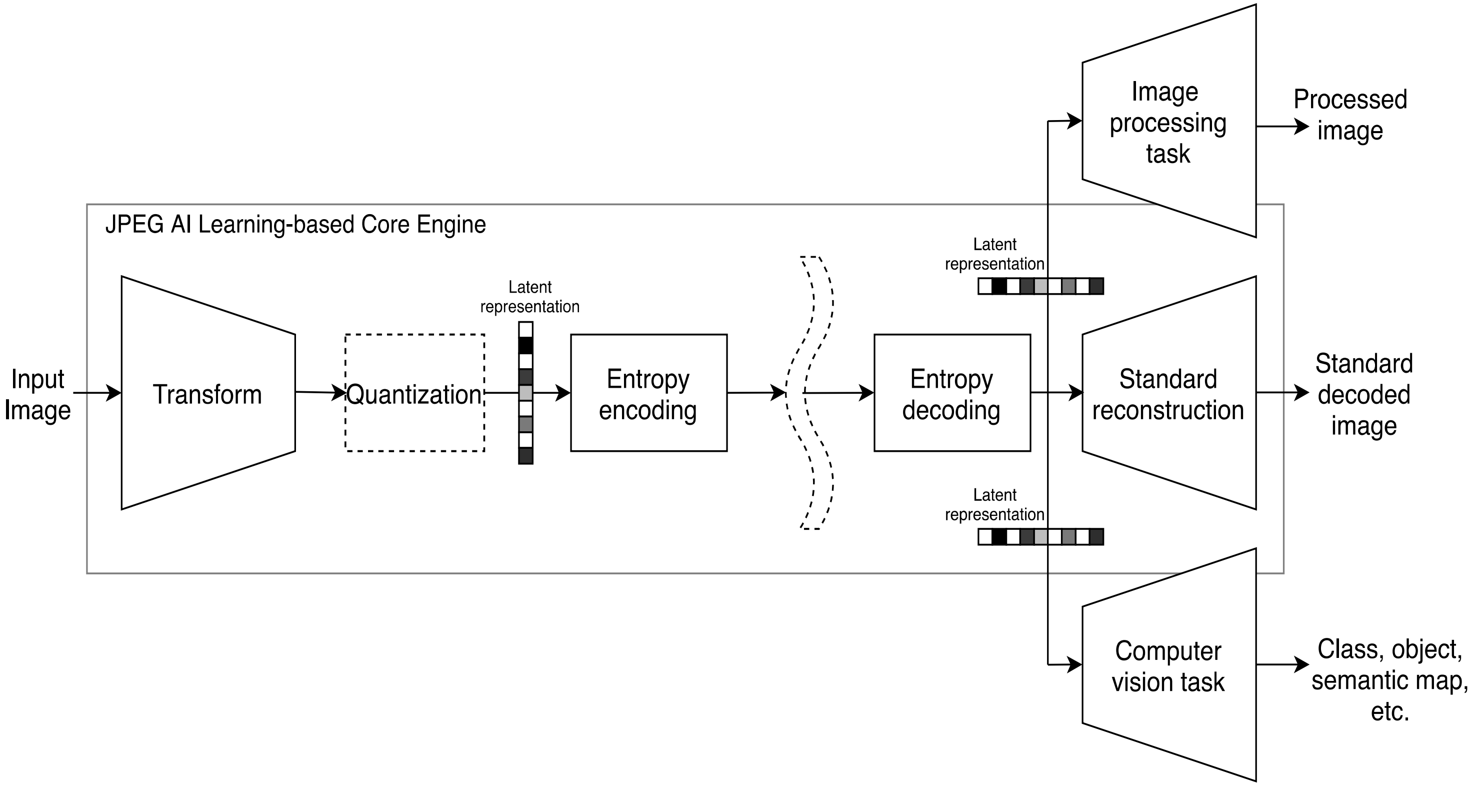
Outline

1. Introduction
2. DL Image Codec Perceptual optimization
3. Subjective Quality Assessment
4. Experimental Results
5. Final Remarks

1 Introduction

Learning-based Image Compression

Objective: Learn a compact representation of images from a large amount of visual data efficiently



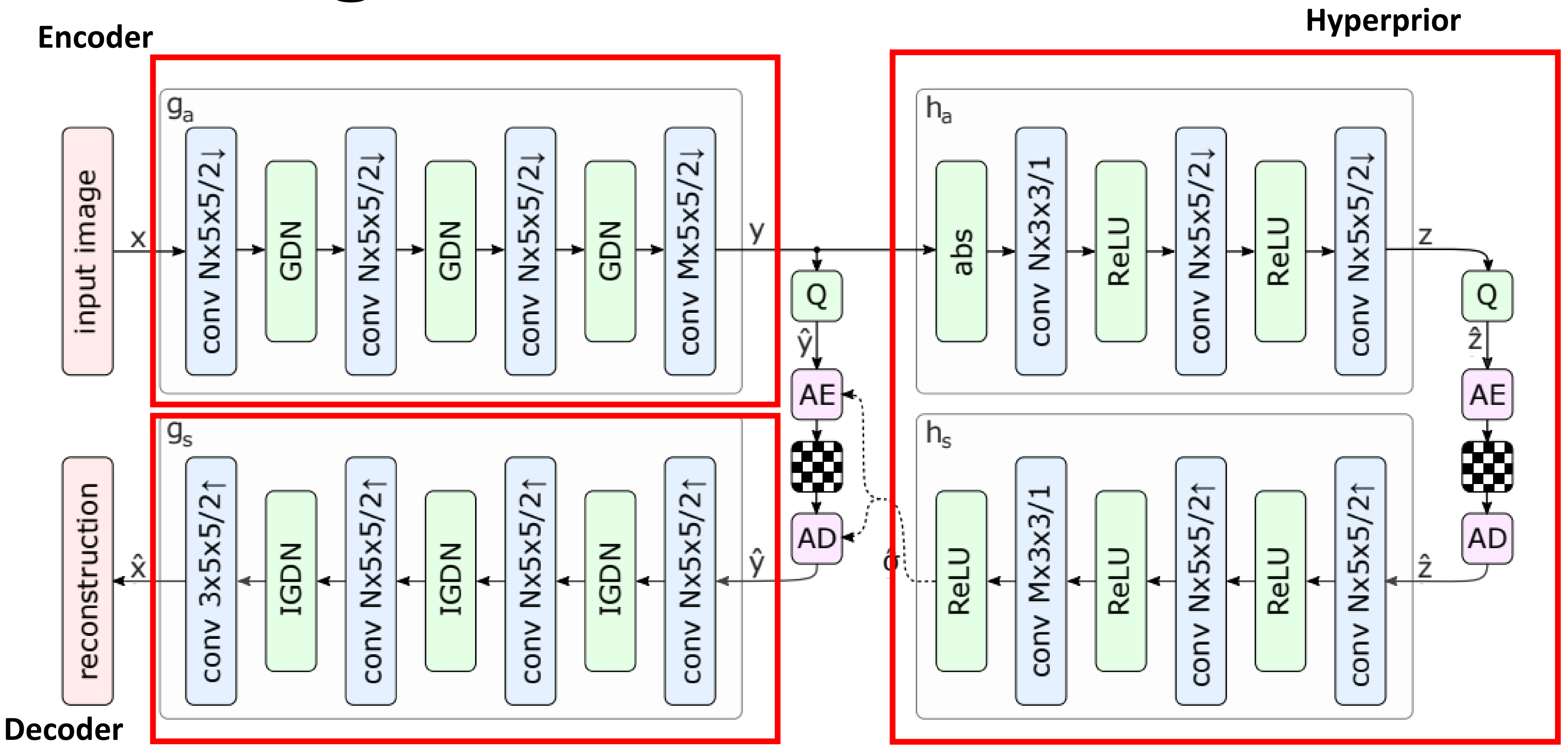
Learning-based vs Conventional codecs

Deep-learning image coding solutions have better compression efficiency than conventional solutions in terms of rate-distortion trade off

JPEG AI CfP Performance Evaluation

TEAMID	BD-rate vs VVC
TEAM14	-32.3%
TEAM24	-29.9%
TEAM16	-17.9%
TEAM12	-3.1%
TEAM22	7.2%
TEAM19	8.6%
TEAM13	10.6%
TEAM21	13.8%
TEAM17	32.0%
TEAM15	51.2%

Learning-based Architecture



Ballé J, Minnen D, Singh S, Hwang SJ, Johnston N. Variational image compression with a scale hyperprior. arXiv preprint arXiv:1802.01436. 2018 Feb 1.

Challenges in Learning-based Image Coding

- Which encoder/decoder architecture is more promising ?
- Which types of processing layers should be used ?
- **Which types of the quality metrics should be used for optimization ?**
  - **Idea: Freeze everything (e.g., the architecture) and only change the image quality metric used in the loss function**

Objective and Contributions

Study the perceptual impact of several image quality metrics for deep learning-based image codec optimization

Subjective assessment campaign which evaluates several learning-based compression models which only differ on the loss function quality metric

2 DL Image Codec Perceptual Optimization

Rate-distortion Tradeoff

- Rate-distortion tradeoff is controlled by λ and expressed by
  - $\mathrm{L} = \lambda\left(D\left(X - \hat{X}\right)\right) + R(\hat{Y}))$
- Quality metric D plays a very important role on the DL-image coding model creation

Metrics	Short Description
MSE	Measures pixel-wise squared differences
SSIM	Measures the degradation in structural information
MS-SSIM	SSIM extension that supports variations in image resolution and viewing conditions
FSIM	Exploits phase congruency and gradient information
GMSD	Measures pixel-wise gradient differences
LPIPS	Measures similarity using deep features
DISTS	Measures structural distortions with a tolerance for texture resampling
NLPD	Measures root mean square error differences in normalized Laplacian domain
VSI	Exploits saliency features for local distortion computation

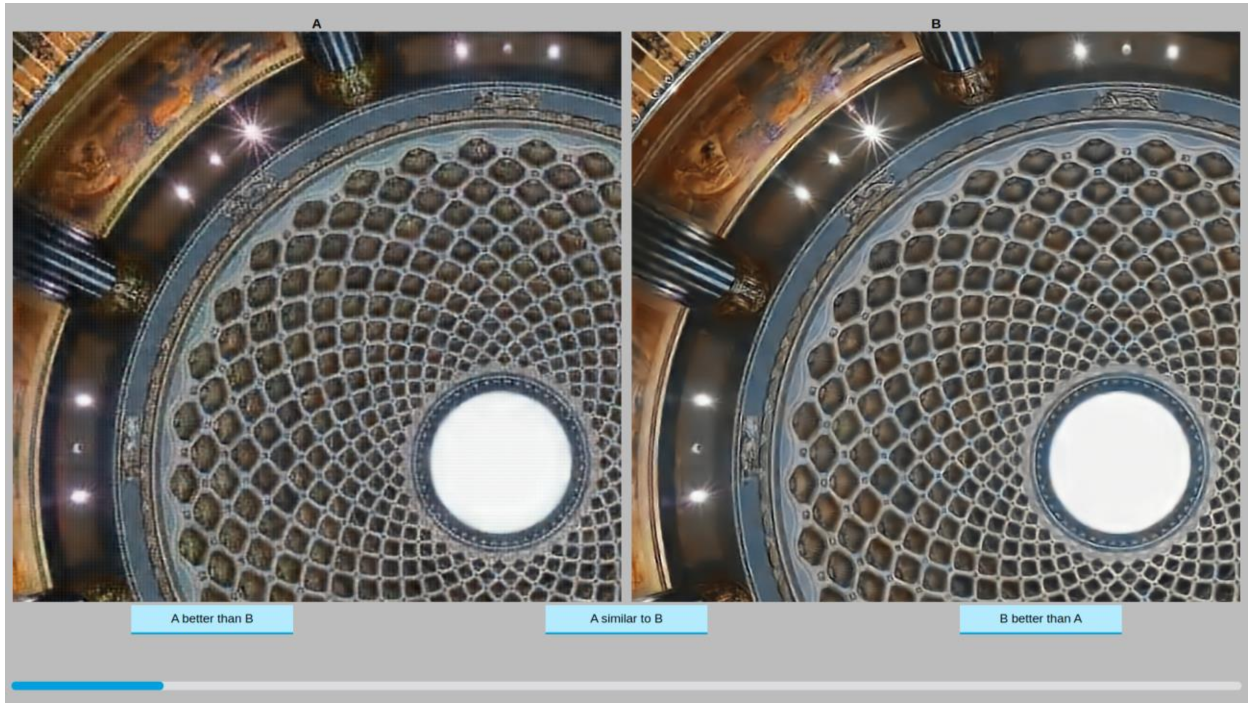
Training Procedure

- Training on patches of JPEG AI training and validation dataset
- Compress AI library implementation of VAE-hyperprior was used
- All the models were pretrained with MSE
- Learning rate = 1e-5
- Image quality metrics in the loss function:
  - DISTS, LPIPS, MSE, MS-SSIM, NLPD, GMSD, FSIM, SSIM, VSI
- 200 epochs for training each model

3 Subjective Quality Assessment

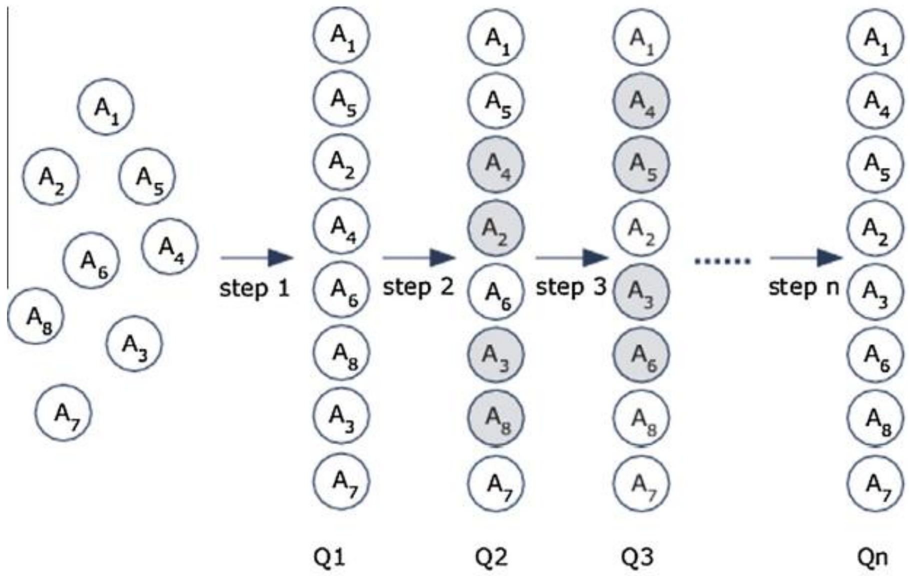
Subjective Evaluation Methodology

- Pairwise comparison (PC) subjective assessment
- Advantages:
  - High accuracy
  - Robustness
  - No training for meaning of the quality scales
- Disadvantages:
  - Long duration
  - Number of pairs for one reference:
    - $\frac{n(n-1)}{2}$



Pairwise Sampling Method

- Objective:
  - Make a shorter and less expensive test
- Iterative approach:
  - One pair never compares twice
  - Only adjacent pairs are compared

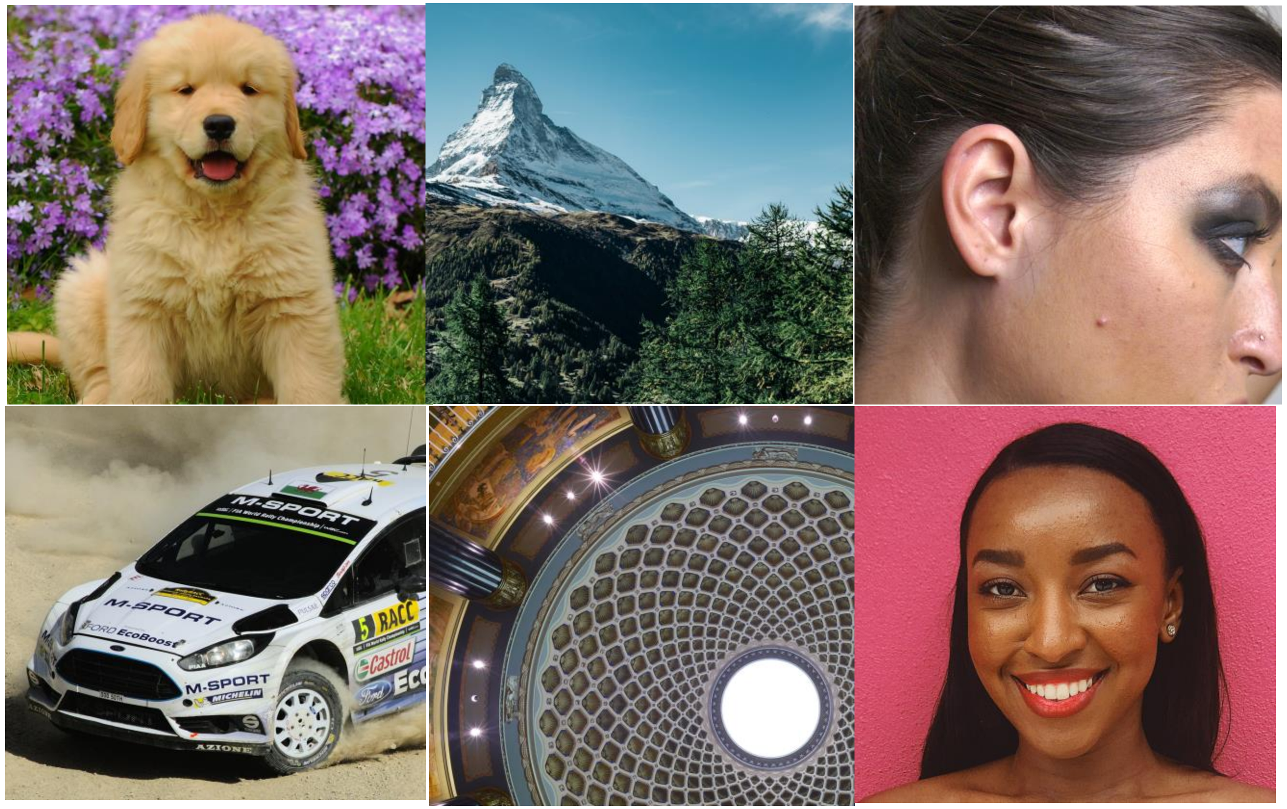


Experimental Setup

- Web-based platform using JavaScript and MongoDB database
- Subjects recruited from Amazon Mechanical Turk (AMT)
- Requirements:
  - Minimum display resolution of 1920×1080
  - Display size must be above 13 inches
- Training phase for the subjects to be familiar with the interface and objective of the test.

Test Material

- Six images of JPEG AI test set
- Images were cropped to fitted side by side layout



Subjective Data Processing

1. Outlier detection
   - Number of transitivity cycles
   - $R = 1 - d/h$

2. Quality score computation
   - PC matrix for each subject is created
   - Group preference matrix is calculated
   - Winning frequencies are inferred
     - Number of votes each metric receives divided by the total number of comparison

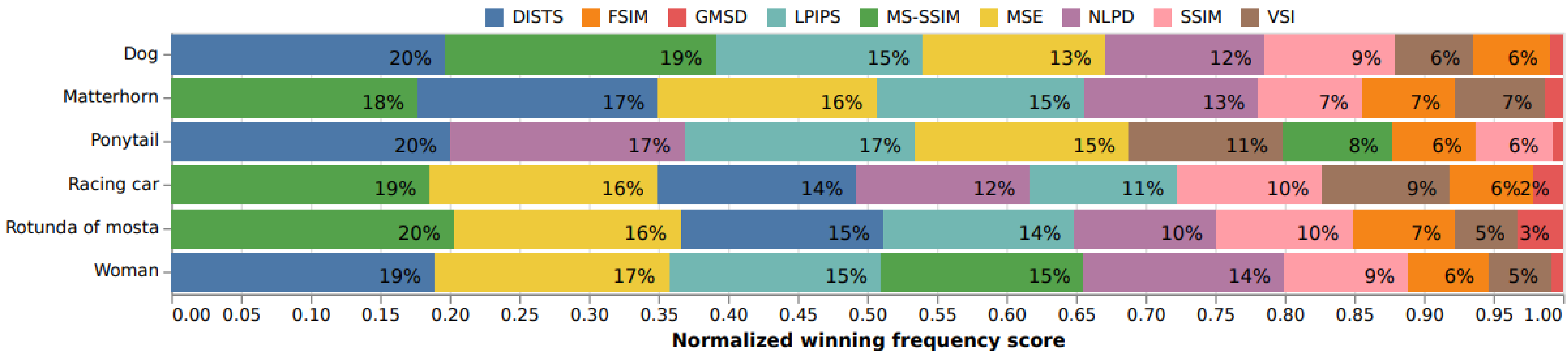
Subjective Test Statistics

- 120 users distributed in three sessions
- Subjects age
  - Between 20 and 60 with average of 34
- Gender distribution
  - 70% male
- Display resolution
  - More common is 1920×1080
- Display size
  - More common is 15 inches
- Number of outliers
  - 2, 4, 6 in three sessions

4 Experimental Results

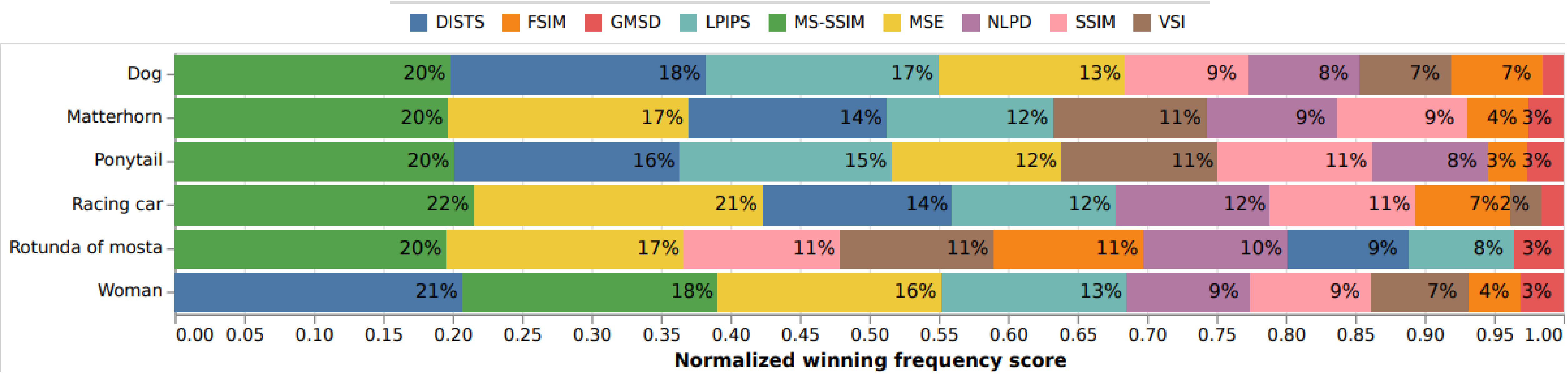
Quantitative Results (Low bitrate)

- Scores were normalized for each test image
- DISTS, MS-SSIM and MSE have the best performance with exception of Ponytail image



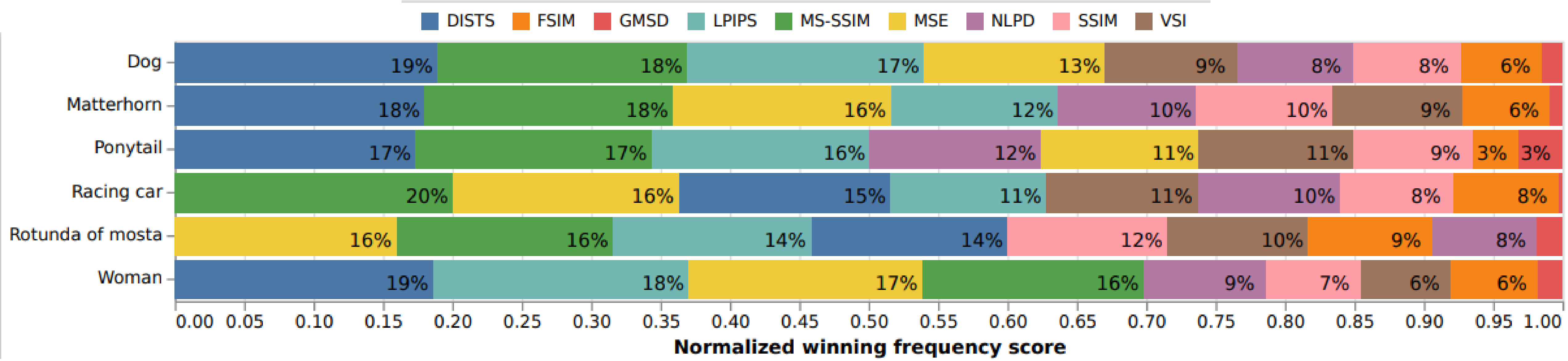
Quantitative Results (Medium bitrate)

- MS-SSIM has the best performance 5 out of six images.
- DISTS has the best performance in Woman image



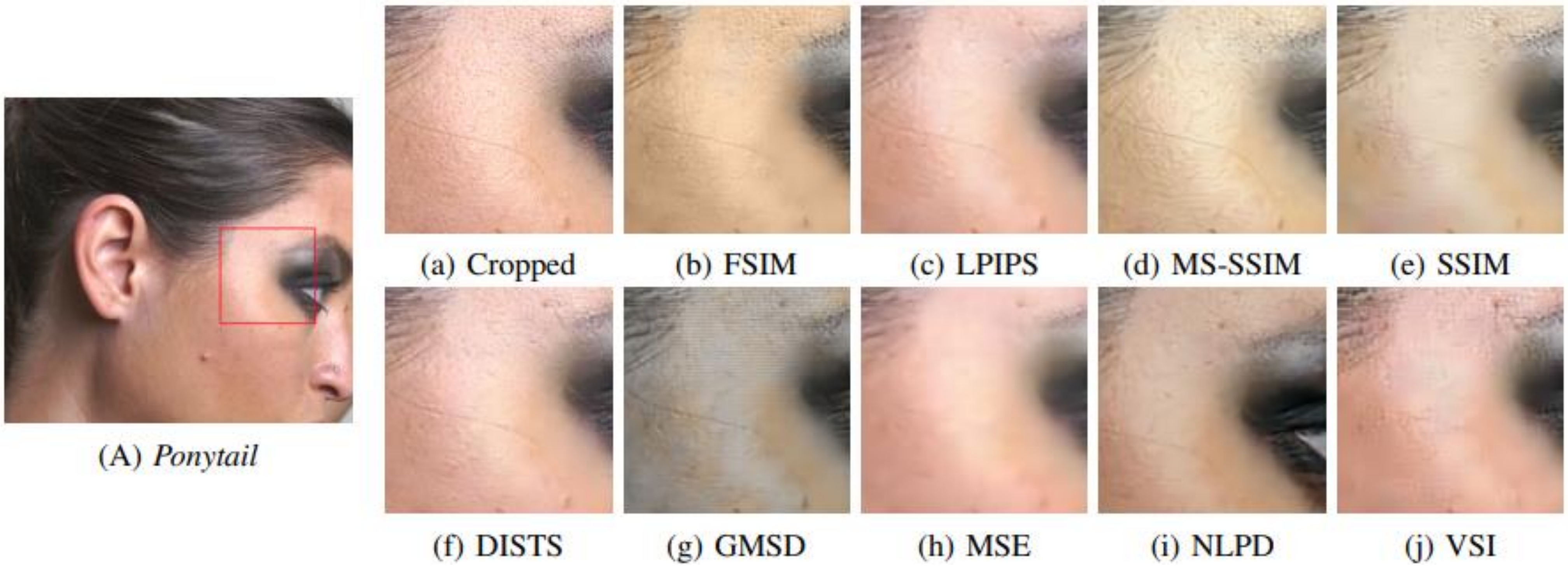
Quantitative Results (High Bitrate)

- DISTS has the highest overall performance except for Rotunda of Most and Racing car where MS-SSIM and MSE provides better performance.
- MSE performs poorly in high bitrates compare to low and medium bitrate.



Qualitative Results

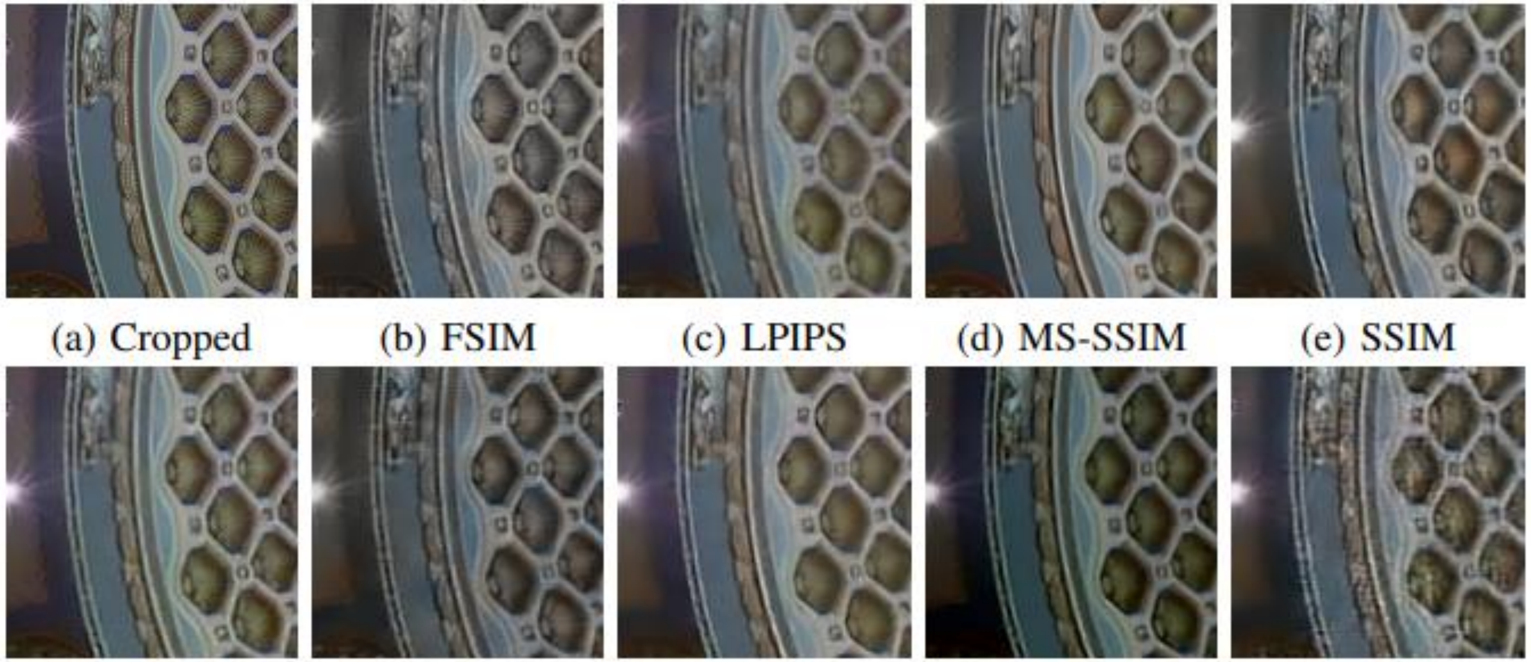
- MS-SSIM model failed to generate the natural skin of the face
- DISTS provides high quality



Qualitative Results

- MS-SSIM model generates more sharp images
- LPIPS and DISTS models generate images with ringing artifacts

(B) *Rotunda Mosta*



(a) Cropped (b) FSIM (c) LPIPS (d) MS-SSIM (e) SSIM

(f) DISTS (g) GMSD (h) MSE (i) NLPD (j) VSI

5 Final Remarks

Final Remarks

- Contributions:
  - Study of the perceptual impact of several image quality metric in the loss function
  - Large scale crowdsourcing pairwise subjective test was performed
- Conclusions:
  - The choice of image quality metric matters !
  - MS-SSIM and DISTS offer the best rate-distortion tradeoff
  - Loss functions better selected for each bitrate could provide performance improvements

Thanks for your attention

For more information email us:
shima.mohammadi@lx.it.pt, joao.ascenso@lx.it.pt