Supplemental Material: Hierarchical Auto-Regressive Model for Image Compression Incorporating Object Saliency and a Deep Perceptual Loss

Yash Patel[2], Srikar Appalaraju[1], R. Manmatha[1]
[1]Amazon
[2]Center for Machine Perception, Czech Technical University, Prague, Czech Republic
patelyas@cmp.felk.cvut.cz, (srikara,manmatha)@amazon.com

This note compliments the material from the main paper: "Hierarchical Auto-Regressive Model for Image Compression Incorporating Object Saliency and a Deep Perceptual Loss" [6]. Due to space and size constraints we had to separate it out.

1. Human Evaluation Setup

The process of obtaining evaluations for human perceptual similarity is detailed in Sec. 2.1 of the main paper [6]. Here in Fig. 1, we provide an illustration from our Mechanical Turk HIT page. Furthermore, Patel et al.[5] details more on Human evaluation study specific to image compression and also why traditional metrics like MSE and MS-SSIM are not the optimal loss functions to optimize for human visual perception.

Figure 1: Sample instance from MTurk HIT. Entire images are shown at the top with the original image in the middle and the image from one method on the left and other on the right. The bottom images are magnified versions of a small window which can be controlled by moving the cursor.

2. Combining Perceptual Similarity Metrics

In Sec. 2.3 of main paper [6], a set of combinations of various learned and hand-crafted metrics has been discussed. The weights for these combinations are learned from the training set of our compression specific perceptual similarity dataset. The details of learning these weights is elaborated in this section.

For instance for a combination of PSNR and LPIPS, let us say the weight given to PSNR is d_1 and that for LPIPS is d_2. Thus the objective is to learn d_1 and d_2. For a sample from the training set:

$$d1 * PSNR(x_i, \hat{x_i}) + d_2 * LPIPS(x_i, \hat{x_i}) = b_i \quad (1)$$

Here b is the 2AFC score for the sample.

Over a set of k randomly selected samples the process is repeated, giving k equations:

$$\begin{bmatrix} d_1 \\ d_2 \end{bmatrix} \begin{bmatrix} PSNR(x_1, \hat{x_1}) & LPIPS(x_1, \hat{x_1}) \\ ... & ... \\ PSNR(x_k, \hat{x_k}) & LPIPS(x_k, \hat{x_k}) \end{bmatrix} = \begin{bmatrix} b_1 \\ ... \\ b_k \end{bmatrix} \quad (2)$$

Eq. 2 can be written in a matrix multiplication form:

$$\mathbf{dA} = \mathbf{b} \quad (3)$$

The least square solution of Eq. 3 is obtained by the pseudo inverse of $\mathbf{A}$. Using SVD decomposition $\mathbf{A} = \mathbf{UDV}^T$, we obtain the combination variables:

$$\mathbf{d} = \mathbf{VD}^{-1}\mathbf{U}^T\mathbf{b} \quad (4)$$

We obtain $\mathbf{d}$ using two samples at a time (4 sets of linear equations) and we perform RANSAC and pick the $\mathbf{d}$ that satisfies the most samples in the train set.

3. Weighted Distortion Losses

Determining the final distortion loss with the saliency mask is detailed in Sec. 3.4. Here, we provide a visual example from the training in Fig. 2.

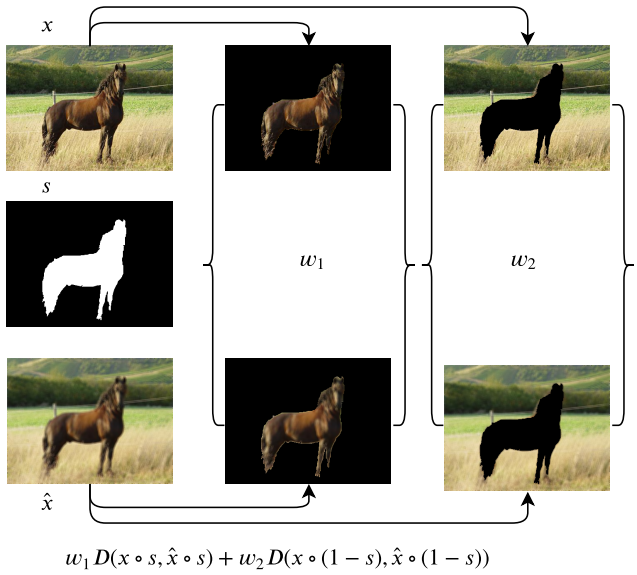

Figure 2: **Weighted Distortion Losses**: The original image $\mathbf{x}$ is decomposed into a salient $\mathbf{x} \odot \mathbf{s}$ - top middle image - and a non-salient $\hat{\mathbf{x}} \odot \mathbf{s}$ - top right image - component. Similarly, the reconstructed image is divided into a salient - bottom middle image - and a non-salient -bottom right image - component. The distortion losses are separately computed on both and are then linearly combined with more weight given to the salient component.

4. Qualitative Results

All the images from the Kodak dataset for our methods and the five competing methods used for human evaluations can be downloaded from this anonymous link click here (420 MB).

In that location you can find all Kodak dataset images at four different bpp's (0.23, 0.37, 0.67 and 1.0) for our method and Competing methods: Mentzer et al [4], Balle et al [1], JPEG2K [7], Lee et al [3] and BPG [2]

The qualitative results are shown in Fig. 3, 4 and 5 where we provide the full images along with selected crops. We also provide all the kodak images compressed from our method at 0.37 bpp in Fig. 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28 and 29.

All the images from the kodak dataset for our methods and the five competing methods used for human evaluations can be downloaded from this anonymous link click here (420 MB).

References

[1] J. Ballé, V. Laparra, and E. P. Simoncelli. End-to-end optimized image compression. *arXiv preprint arXiv:1611.01704*, 2016.

[2] F. Bellard. Bpg image format, 2014.

[3] J. Lee, S. Cho, and S.-K. Beack. Context-adaptive entropy model for end-to-end optimized image compression. *ICLR*, 2019.

[4] F. Mentzer, E. Agustsson, M. Tschannen, R. Timofte, and L. Van Gool. Conditional probability models for deep image compression. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4394–4402, 2018.

[5] Y. Patel, S. Appalaraju, and R. Manmatha. Human perceptual evaluations for image compression. *arXiv preprint arXiv:1908.04187*, 2019.

[6] Y. Patel, S. Appalaraju, and R. Manmatha. Hierarchical auto-regressive model for image compression incorporating object saliency and a deep perceptual loss. *arXiv preprint arXiv:2002.04988*, 2020.

[7] A. Skodras, C. Christopoulos, and T. Ebrahimi. The jpeg 2000 still image compression standard. *IEEE Signal processing magazine*, 2001.

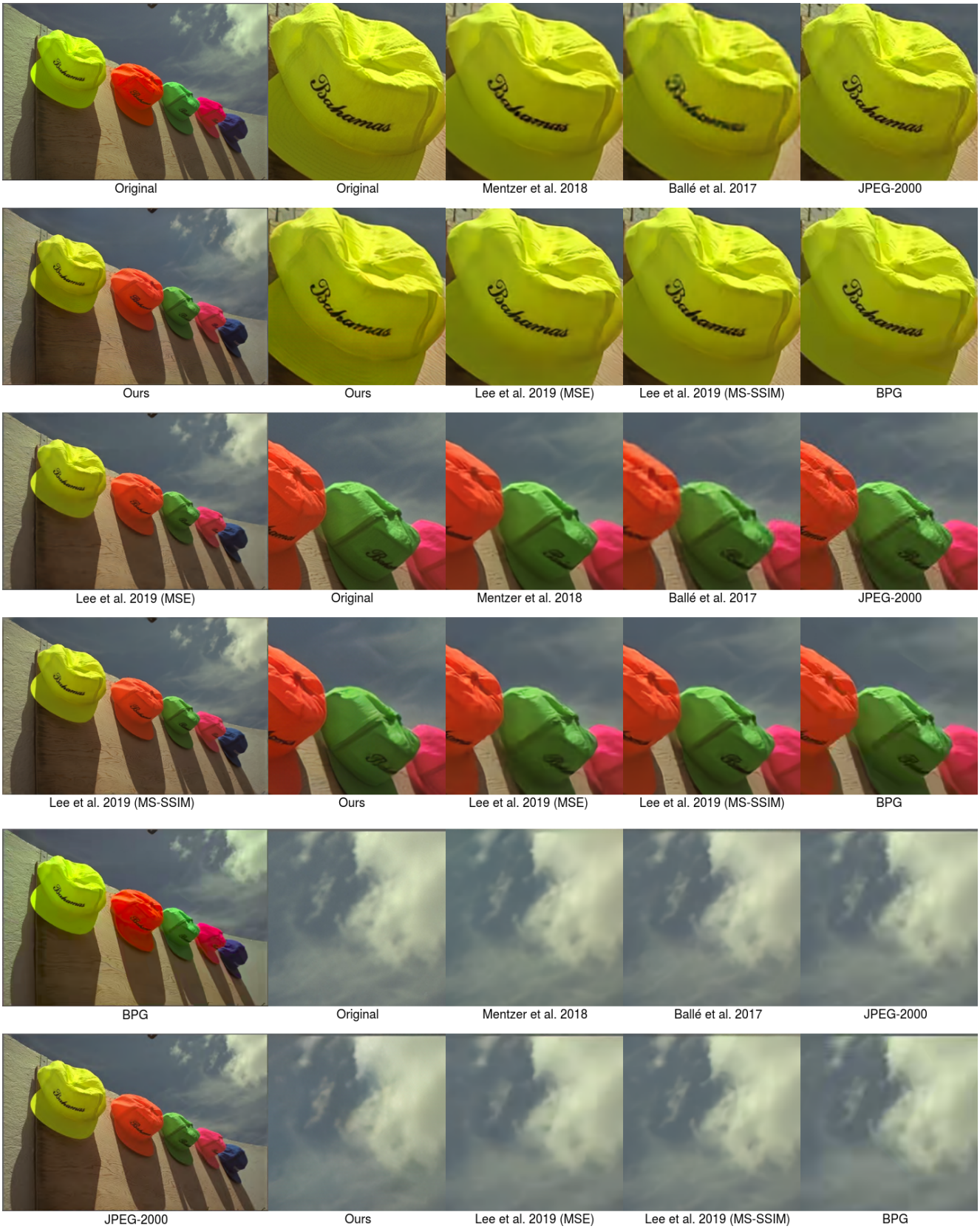

Figure 3: Kodak image kodim03.png at **0.37** bpp. Compared across methods: [4, 1, 7, 6, 3, 2]

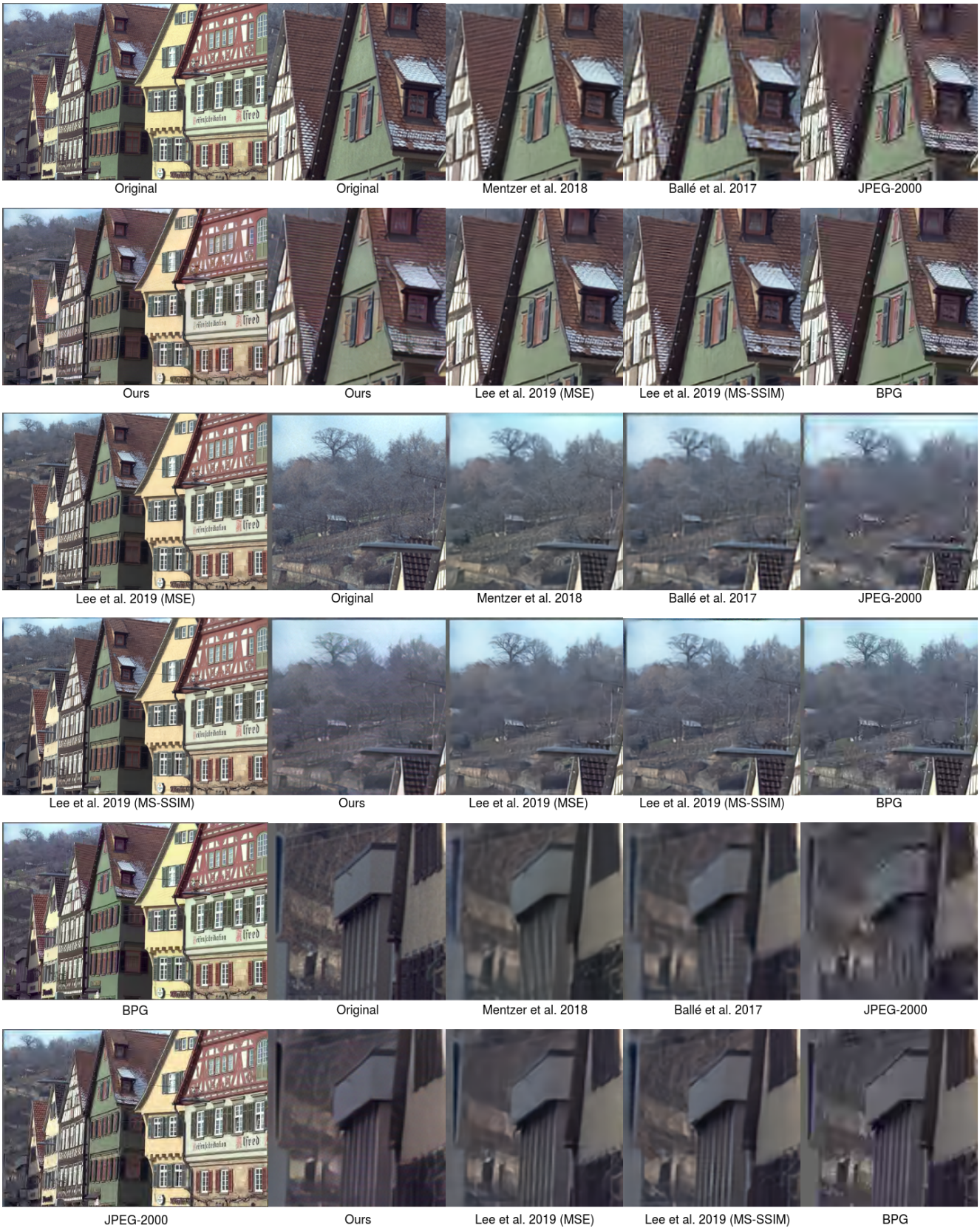

Figure 4: Kodak image kodim08.png at **0.37** bpp.

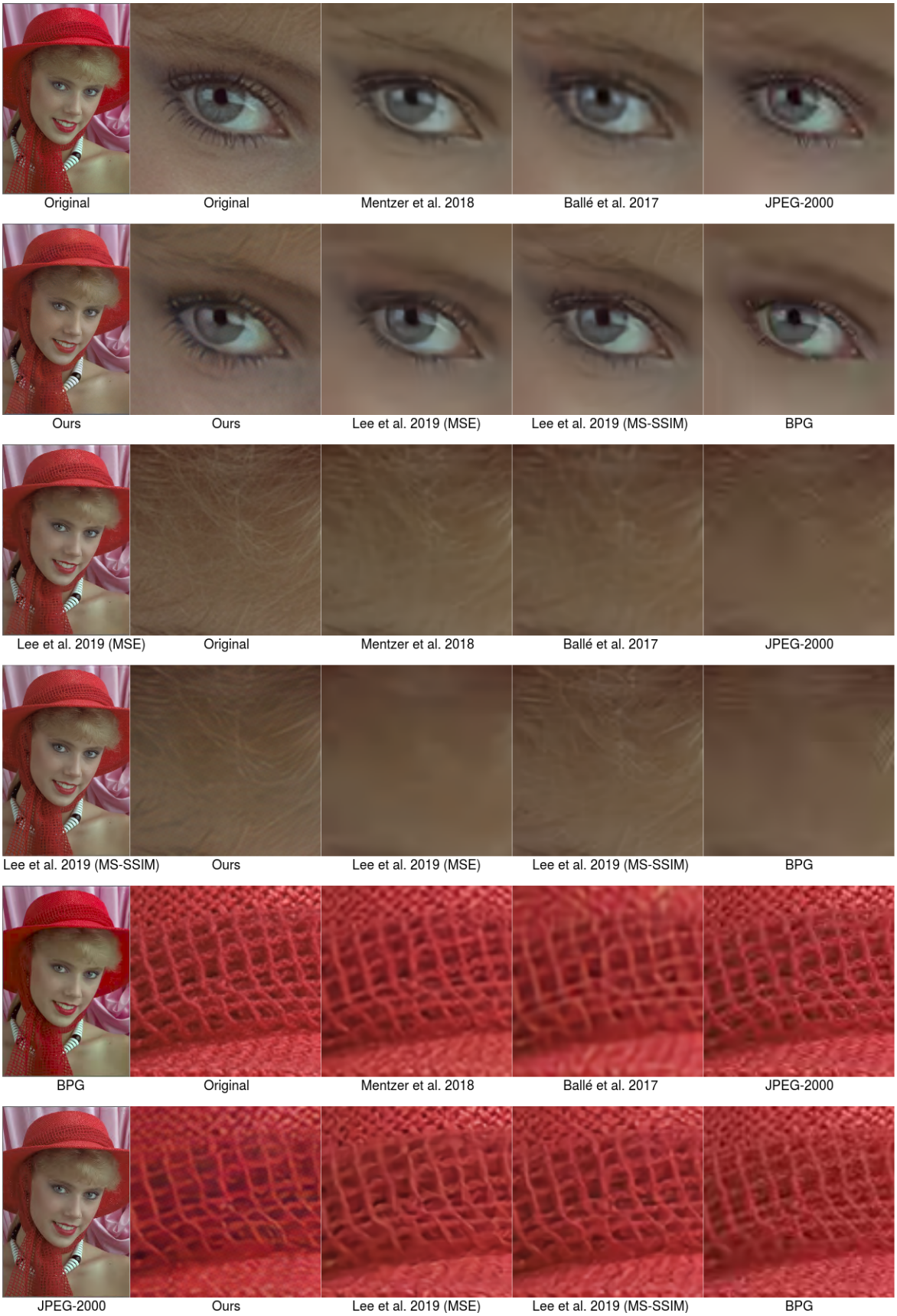

Figure 5: Kodak image kodim04.png at **0.22** bpp.

Figure 6: Kodim01.png at 0.37 bpp

Figure 7: Kodim02.png at 0.37 bpp

Figure 8: Kodim03.png at 0.37 bpp

Figure 9: Kodim04.png at 0.37 bpp

Figure 10: Kodim05.png at 0.37 bpp

Figure 11: Kodim06.png at 0.37 bpp

Figure 12: Kodim07.png at 0.37 bpp

Figure 13: Kodim08.png at 0.37 bpp

Figure 14: Kodim09.png at 0.37 bpp

Figure 15: Kodim10.png at 0.37 bpp

Figure 16: Kodim11.png at 0.37 bpp

Figure 17: Kodim12.png at 0.37 bpp

Figure 18: Kodim13.png at 0.37 bpp

Figure 19: Kodim14.png at 0.37 bpp

Figure 20: Kodim15.png at 0.37 bpp

Figure 21: Kodim16.png at 0.37 bpp

Figure 22: Kodim17.png at 0.37 bpp

Figure 23: Kodim18.png at 0.37 bpp

Figure 24: Kodim19.png at 0.37 bpp

Figure 25: Kodim20.png at 0.37 bpp

Figure 26: Kodim21.png at 0.37 bpp

Figure 27: Kodim22.png at 0.37 bpp

Figure 28: Kodim23.png at 0.37 bpp

Figure 29: Kodim24.png at 0.37 bpp