Benchmarking JPEG XL image compression

Jyrki Alakuijala[a], Sami Boukortt[a], Touradj Ebrahimi[c], Evgenii Kliuchnikov[a], Jon Sneyers[b],
Evgeniy Upenik[c], Lode Vandevenne[a], Luca Versari[a], Jan Wassenberg[1*a]
[a] Google Switzerland, 8002 Zurich, Switzerland; [b] Cloudinary, Santa Clara, CA 95051, USA;
[c] Multimedia Signal Processing Group, École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland.

ABSTRACT

JPEG XL is a practical, royalty-free codec for scalable web distribution and efficient compression of high-quality photographs. It also includes previews, progressiveness, animation, transparency, high dynamic range, wide color gamut, and high bit depth.

Users expect faithful reproductions of ever higher-resolution images. Experiments performed during standardization have shown the feasibility of economical storage without perceptible quality loss, lossless transcoding of existing JPEG, and fast software encoders and decoders. We analyse the results of subjective and objective evaluations versus existing codecs including HEVC and JPEG.

New image codecs have to co-exist with their previous generations for several years. JPEG XL is unique in providing value for both existing JPEGs as well as new users. It includes coding tools to reduce the transmission and storage costs of JPEG by 22% while allowing byte-for-byte exact reconstruction of the original JPEG. Avoiding transcoding and additional artifacts helps to preserve our digital heritage.

Applications require fast and low-power decoding. JPEG XL was designed to benefit from multicore and SIMD, and actually decodes faster than JPEG. We report the resulting speeds on ARM and x86 CPUs. To enable reproduction of these results, we open sourced the JPEG XL software in 2019.

Keywords: image compression, JPEG XL, perceptual visual quality, image fidelity

1. INTRODUCTION

Image compression nowadays is universally employed worldwide in every aspect of human lives: from consumer hand-held electronic devices equipped with high-resolution cameras and screens to heavy industry and space exploration. The research in this field continues to be a hot topic, even though it has been investigated for over three decades. The experts in multimedia signal processing define two main approaches in compression: lossless and lossy. The former approach guarantees to preserve the information in such a way that the reconstructed image corresponds bit-exactly to the digital representation of the original image being compressed. The latter approach takes into account the properties of the human visual system, the intermediate medium, and the environment in order to identify parts of visual information that can be discarded. The concept of perceptual visual quality is introduced in order to optimize the performance of such compression.

When the scientific community achieves a significant break-through, the technology is transferred to practical applications. These applications are often subject to standardization: a necessary phase that is required for adoption of a

---

[1*]janwas@google.com

technology by the global industry. JPEG XL is a technology that is currently undergoing the latest steps of its standardization conducted by members of JPEG (Joint Photographic Experts Group) formally known as the Working Group 1 (WG1) "Coding of Still Pictures" of Subcommittee 29 (SC29) "Coding of Audio, Picture, Multimedia and Hypermedia Information" of Joint Technical Committee 1 (JTC1) between the International Organization for Standardization (ISO), the International Electrotechnical Commission (IEC), and the International Telecommunication Union (ITU).

When images are compressed using lossy compression and then reconstructed, the resulting signals differ from the originals, and the amount of information loss is typically proportional to the level of compression. Hence, one needs to assess the perceptual visual quality with respect to the compression ratio or simply the bitrate of the compressed bitstream in order to find the optimal parameters or a trade-off between the quality and the size. Methods for automatically predicting perceptual visual quality have also been devised. Objective metrics are typically used when it is impossible or infeasible to perform subjective evaluation experiments, such as in a rate-distortion optimization feedback loop.

In this paper, we benchmark the novel JPEG XL codec against popular modern methods for image compression. We compare the results of subjective and objective perceptual quality evaluations performed on a dataset of images compressed at various bitrates. Additionally, we assess the encoding and decoding speed of the codecs under investigation when executed with different levels of parallelization.

The remainder of this paper is organized as follows: Section 2 provides a very short overview of the research related to the current study which includes evaluations of the performance of earlier codecs. Section 3 briefly describes the architecture and the operating modes of the novel JPEG XL codec. Section 4 covers the creation of the dataset used in the subjective and objective evaluation experiments and the selection of the reference codecs. Section 5 presents the results of the subjective evaluation of perceptual visual quality consisting in the mean opinion scores, followed by a short discussion. Section 6 lists numerical results of the objective perceptual visual quality measurements. Section 7 addresses the performance of JPEG XL in terms of the encoding speed compared to the reference codecs. Finally, Section 8 concludes the paper.

2. RELATED WORK

Due to the fact that JPEG XL is a very new image compression codec, to the best of our knowledge, there has not been any formal benchmarking or evaluation of its current performance. We feel obliged, however, to refer to the similar work that was performed with regard to earlier codecs at the time when they were introduced. This short overview will also help to identify the place of JPEG XL in the universe of image compression.

The first widely known attempt to improve still image compression, after JPEG[1] had gained great success[2], was JPEG 2000[3,4]. Many studies have been conducted in order to assess the performance of JPEG 2000. De Simone *et al.* provide the results of objective performance evaluation[5] which include JPEG and JPEG 2000. The authors compute PSNR and MS-SSIM metrics on a dataset of compressed images based on ten different contents: JPEG 2000 outperforms JPEG in all cases. Another similar work[6] by De Simone *et al.* includes VIFp metric and supports the same hypothesis for 15 different contents. Subjective perceptual quality evaluation which includes JPEG 2000 and was performed in a controlled environment[7] shows that JPEG 2000 is significantly superior to JPEG for the bitrates lower than 0.75 bpp in all cases with 95% confidence. Other studies support the same conclusions[8,9].

Further in this paper, we use HEVC Intra codec[10,11] as one of the anchors. A study on subjective and objective evaluation of HEVC Intra coding was presented in the paper[12] by Hanhart *et al.*, where the rate-distortion plots of PSNR against the bitrate show that HEVC is better than JPEG in all the cases and is better than JPEG 2000 in five cases out of six. The authors also provide the results of subjective evaluation, which show with 95% confidence that HEVC is significantly better than JPEG in five cases out of six for bitrates lower than 1.0 bpp. However, JPEG 2000 shows performance

similar to HEVC Intra except for very low bitrates where HEVC is better in four cases out of six. In addition, another subjective evaluation study[13] by Hanhart *et al.* shows the superiority of HEVC over AVC for video.

3. JPEG XL

JPEG XL is designed for a wide range of use cases centered around capturing, storing, and transmitting photographic images, graphics, and mixed content (e.g. screenshots) as well as animations/bursts. It supports HDR, alpha channels, layers, thumbnails, lossless and progressive coding. To provide benefit to existing producers and consumers of JPEG, it also allows lossless transcoding and enhanced decoding of existing JPEG images.

Those features are delivered by three separate coding modes: a dedicated predictor and entropy coder for JPEG, a "modular" mode suitable for lossless and/or lossy encoding of synthetic images, and the main mode optimized for human perception. The latter consists of a psychovisual color space, variable-sized DCT with chroma from luma, and ANS-based entropy coding.

The coding tools enable a balance between three criteria:

- high fidelity to the source image,
- fast encoding and/or decoding even in software,
- competitive compression ratio (typically 20:1 to 50:1).

The three coding modes are shown in the encoder block diagram in Figure 1.
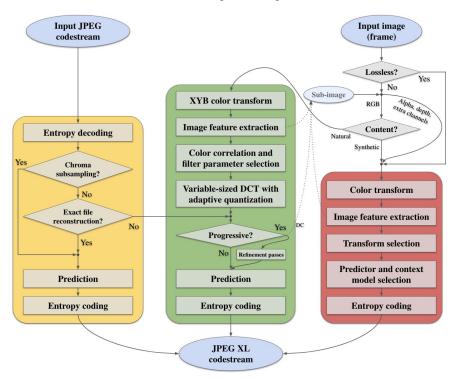

Fig. 1. Block diagram of the JPEG XL encoder, showing the lossless JPEG transcoding, perceptually-optimized and modular modes (left to right).

This paper focuses on the DCT mode (green block), which is generally used for lossy compression of photographic material (i.e. all but the dice and screenshot images described below). It is an instance of the traditional block-based

coding model, in which the components have been upgraded with respect to JPEG. The "XYB color transform" is inspired by properties of the human visual system and, in particular, separates the blue channel from the other chroma information. "Image feature extraction" denotes detection of splines, patches (similar to Intra block copy) and noise or grain for synthesis in the decoder. "Color correlation" is also known as chroma from luma, which serves to decorrelate the blue channel. Unlike JPEG, the DCT in JPEG XL supports various square and rectangular block sizes from 2x2 to 32x32, signaled as one identifier per 8x8 pixels. "Prediction" encompasses an adaptive (history-based) predictor for DC. Finally, "Entropy coding" is either Huffman or alias rANS with context modeling.

The JPEG XL reference software is available under a free and open source software license (Apache 2.0), which includes a royalty-free patent grant from the contributors: http://gitlab.com/wg1/jpeg-xl.

4. DATASET AND SOFTWARE

In order to benchmark the proposed JPEG XL codec, we need to select the reference (or anchor) compression methods, their implementations, and encoding parameters, and define a set of original images that will be used to create the stimuli. The selection of anchors is based on the general popularity of the codecs and the availability of the software implementations. We only intended to use codecs which are standardized internationally.

The list of codecs selected for the benchmarking is: (a) JPEG, (b) JPEG 2000, (c) HEVC.

JPEG[1] is the most widely used image compression. Despite the fact that it does not show the best performance in terms of compression efficiency, we must consider it in our study due to its universal adoption. In order to encode the images with JPEG, we used two different implementations: Guetzli v1.0.1 for higher bitrates and Libjpeg v.1.56 for lower bitrates. In Figure 2, one can find in which exact cases we used the former or the latter. We used two different implementations with an assumption that Guetzli produces higher quality results. However, due to its design, it cannot produce results at very low bitrates.

JPEG 2000[3,4] is another international standard which is very well adopted in particular niche industries such as content creation and production. For the purpose of our study, we used the version of JPEG 2000 with visual optimisation[14]. We used Kakadu v7.10.2 software to encode the images.

HEVC[10] is one of the latest internationally standardized codecs for video and images. In our work, we used HEVC Intra Coding[11] in YUV444 color space to compress images. HM-16.20+SCM-8.8 was used to encode files. Due to the fact that HEVC only takes as an input images represented in YCbCr color space, the original RGB images were reversibly converted to YUV444 according to the ITU recommendation[15(p709)] and the reconstructed images were converted back to RGB before any comparison. Note that this particular encoder for HEVC is, to the best of our knowledge, not commonly used in practice, perhaps due to its low encoding throughput. For this reason, we also include the x265 software in the speed measurements in Section 7.

The actual bitrates of the compressed images can be found in Figure 2. A more detailed description of the dataset which includes the command line parameters for all the codecs can be found in a report contributed to the JPEG Committee[16].

The dataset of stimuli used in subjective and objective perceptual quality evaluations is based on eight original images depicted in Figure 3. The original contents include six photographic and two non-photographic pictures of different sizes specified in Table 4.
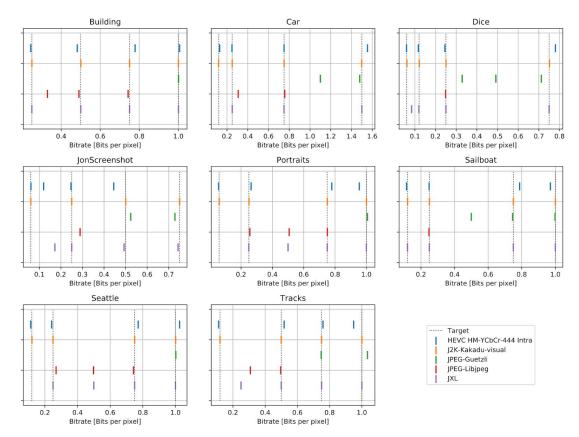
Fig. 2. Bitrates of the stimuli included in the dataset.

Fig. 3. Original images used in the subjective and objective evaluation of perceptual visual quality.

Tab. 4. Description of the original images used to create the evaluation dataset.

N	File	Type	Name	Resolution	Bit Depth	Color Format
1	nonphoto_Dice_800x600_8bit_sRGB.png	nonphoto	Dice	800x600	8bit	sRGB
2	nonphoto_JonScreenshot_1200x1400_8bit_sRGB.png	nonphoto	JonScreenshot	1200x1400	8bit	sRGB
3	photo_Building_938x1400_8bit_sRGB.png	photo	Building	938x1400	8bit	sRGB
4	photo_Car_1016x676_8bit_sRGB.png	photo	Car	1016x676	8bit	sRGB
5	photo_Portraits_1134x756_8bit_sRGB.png	photo	Portraits	1134x756	8bit	sRGB
6	photo_Sailboat_1200x866_8bit_sRGB.png	photo	Sailboat	1200x866	8bit	sRGB
7	photo_Seattle_1072x712_8bit_sRGB.png	photo	Seattle	1072x712	8bit	sRGB
8	photo_Tracks_1200x866_8bit_sRGB.png	photo	Tracks	1200x866	8bit	sRGB

5. SUBJECTIVE EVALUATION

In this section, we provide the results of the subjective evaluation of perceptual visual quality performed on the dataset described above. The experiment was conducted in the test laboratory of Multimedia Signal Processing Group at EPFL. We describe the methodology used in the evaluations and follow with a discussion.

5.1 Methodology

The experiment was conducted in the quality test laboratory of Multimedia Signal Processing Group at EPFL, which complies with the ITU recommendations on subjective evaluation of perceptual visual quality described in the document[17(p500)]. Such a controlled experimental environment intends to assure the reproducibility and the reliability of the results by eliminating possible unaccounted factors which could affect the subjects' opinion.

The monitor used in the experiment was EIZO CG301W. It has a resolution of 2560x1600 with a pixel density of 101 ppi and an 8-bit panel. The monitor was calibrated in compliance with the ITU recommendations[18] according to the following profile: sRGB Gamut, D65 chromaticity of the background, 80 cd/m^2 of maximum brightness. The ambient illumination in the room was set to a level that can be described as low compared to the brightness of the screen as it is recommended in Annex 1 of the document[18].

In order to evaluate the subjective visual quality, we followed the degradation category rating (DCR-HR) method described in the ITU recommendations[19] with hidden reference and simultaneous presentation. The stimuli were presented side-by-side with one of the pictures being the reference and the other an assessed image. The subjects were asked to grade the impairment using a five-level scale: 5 - imperceptible; 4 - perceptible, but not annoying; 3 - slightly annoying; 2 - annoying; 1 - very annoying. In the beginning of each session three training examples of "very annoying", "slightly annoying", and "imperceptible" quality were shown to subjects in order to make them familiar with the procedure and the quality levels.

27 subjects participated in the experiment. One of them was identified as an outlier (as referenced in the next section) and the corresponding data was eliminated. The remaining subjects were represented by 11 females and 15 males of ages ranging from 21 to 35 years old with the mean of 28.08 and the median of 29.00. All the participants passed the visual acuity examination and the Ishihara test for color perception.

5.2 Results of subjective evaluation

The raw data obtained from the subjective evaluation experiment were analysed statistically. Firstly, we performed the outlier detection according to the ITU recommendations[17(p500)] which resulted in eliminating the scores from one subject. Afterwards, the mean opinion scores (MOS) and the corresponding 95% confidence intervals (CI) were computed for each stimulus. The results of this analysis can be found in Figure 5.
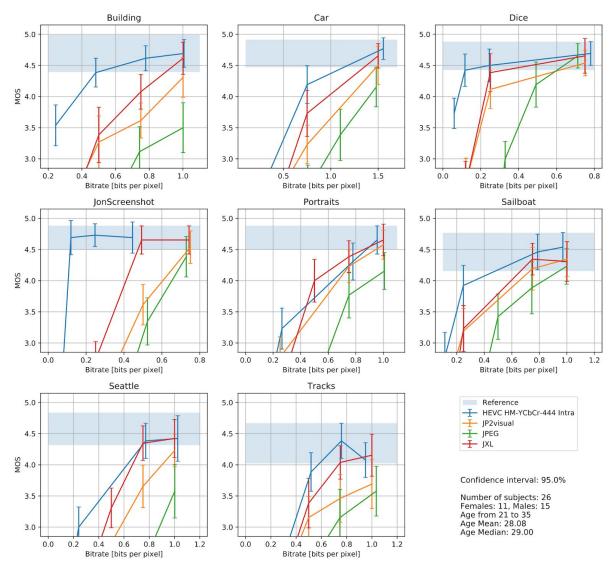
Fig. 5. Subjective mean opinion scores (MOS) with 95% confidence intervals (CI) plotted against the bitrates, for MOS > 3.0 only. The area filled with light-blue color represents the confidence interval of the hidden reference.

5.3 Discussion

Note that the laboratory and evaluation conditions differ from typical home or office usage in two ways. First, the DCR-HR methodology is particularly suitable for quality assessments of lower quality stimuli; in this experiment, most pictures are compressed to between 0.25 and 1.0 bpp and tend to be of higher quality. This is much lower than the mean bitrates of 2 to 3 observed for internet JPEGs, as shown in Figure 6. The preferred operating point of JPEG XL is a bitrate of 1 to 1.5, typically sufficient for visual transparency. Only a few of the rate points in the subjective evaluation fall into this range.
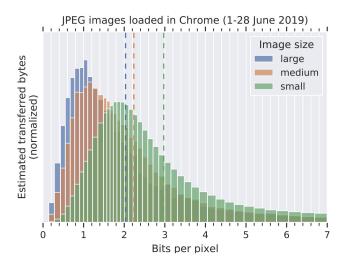
Fig. 6.[20] Number of JPEG images downloaded at various bitrates, times that bitrate. Source: billions of samples from pseudonymous, opt-out Chrome browser telemetry. The results are similar for mobile and desktop. Small indicates 100-400 pixels per side and large denotes ≥ 1000 pixels.

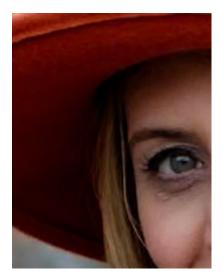Second, the screen brightness of 80 cd/m^2 is darker than the more typical 200 cd/m^2. Although the significance of this is unknown, the JPEG XL images were optimized for 250 cd/m^2, leading to excess rate allocation in dark areas. Conversely, artifacts from other codecs in the dark regions were presumably harder to detect.

Despite these constraints, the evaluation suggests two interesting trends. First, HEVC with SCC generally results in better rate/distortion performance when compared to JPEG XL at low bitrates (< 0.75), better preserving sharp lines and smooth areas (example in Figure 7). Note that our browser telemetry indicates bitrates below 0.75 are less common in internet use. Conversely, JPEG XL is better able to preserve texture and fine details at higher bitrates. HEVC at times requires higher bitrates to preserve details, whereas JPEG XL is more predictably visually lossless at its default setting (example in Figure 8). Note that this image was not included in the subjective evaluation, but we believe it to be representative of this issue.

Fig. 7. JPEG XL (left) and HEVC (right) at about 0.6 bits per pixel. Note grainy artifacts on the left and missing lines on the right.
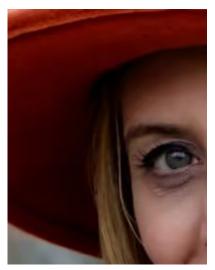
Fig. 8. JPEG XL (left), original (middle) and HEVC (right) at about 1.4 bits per pixel. Note the loss of detail in the right image below the eye and especially on the hat, even at this relatively high bitrate.

6.   OBJECTIVE QUALITY METRICS

In this section, we briefly introduce two objective quality metrics and present the results. We characterize perceptual quality using the "Butteraugli" measure and also include MS-SSIM as a reference point because it is widely used.

6.1 Dataset and software

The 10 input images are the six "photo" images from Section 4 plus the remaining four photographic images from a set curated by JPEG. The image sizes range from 0.6 to 1.3 megapixels. Overall bitrates are defined as the geomean of per-image bitrates. The codecs and Butteraugli versions are specified in Section 7.1. MS-SSIM was computed using libvmaf at revision 55001aa (March 3rd, 2020).

6.1.1   JPEG XL

Bitrates 0.750 bpp, 1.000 bpp, 1.500 bpp and 2.000 bpp were achieved with the following command:

```
tools/benchmark_xl --input='.../photo_*.png' \
    --codec jxl:d2.785,jxl:d1.876,jxl:d1.023,jxl:d0.649
```

Note that these encoder settings (defaulting to "squirrel" speed) do not optimize for the Butteraugli metric.

6.1.2   JPEG (libjpeg-turbo)

Bitrates 0.749 bpp, 0.991 bpp, 1.544 bpp and 2.038 bpp were achieved with the following command:

```
tools/benchmark_xl --input='.../photo_*.png' \
    --codec jpeg:q39,jpeg:q57,jpeg:q78,jpeg:q86
```

6.1.3   JPEG (libjpeg)

Bitrates 0.753 bpp, 0.995 bpp, 1.485 bpp and 1.970 bpp were reached using qualities 29, 44, 69 and 80.

6.1.4   JPEG 2000 (JP2visual)

We used visually-optimized[14] JPEG 2000 outputs generated as reported[16] after confirming that they lead to better metric results than the MSE-optimized mode of Kakadu, as expected.

6.1.5   HEVC-HM-YUV444

Bitrates 0.728 bpp, 0.989 bpp, 1.438 bpp and 2.000 bpp were reached with QP 26, 23, 19 and 15.

6.1.6   HEVC-x265-YUV444

Bitrates 0.737 bpp, 1.000 bpp, 1.443 bpp and 2.035 bpp were achieved with QP 30, QP 27, QP 23 and QP 19.

6.2 Butteraugli

Butteraugli is a full-reference image quality assessment metric under development at Google since 2015. It models several observed properties of the human visual system:

- Eye jitter and contrast sensitivity via Gaussians and a Laplacian pyramid;
- Visual masking using a function of local absolute differences in 2 chroma channels;
- Sensitivity to line-like structures by convolution with rotated line-shaped kernels;
- Multiscale behavior by computing the metric on a subsampled input.

Input images are converted to a psychovisual color space based on LMS with a biased-log transfer function. This is similar but not identical to the color space used in JPEG XL, which uses a cube-root transfer function. The per-pixel error map is summarized to a quality measure using the max norm (at high qualities) or "p-norm", the average of the 3, 6 and 12-norms. The overall result is the geomean of the per-image norms.

In our experience, Butteraugli is a useful proxy for image quality, especially around small multiples of one "Just-noticeable difference". Unpublished independent results have shown it correlates well with subjective assessments, even at wider quality ranges, for distortions caused by classical (non-deep-learned) codecs. In future work, we plan to describe Butteraugli in detail and characterize its performance.

Note that the default encoder of JPEG XL (named "squirrel", faster animals indicate higher encode throughput) and all faster ones (including "cheetah" used in Section 7) **do not optimize for Butteraugli**.

Figure 9 shows the Butteraugli max norm results at the target bitrates of 0.75, 1, 1.5, and 2 bpp for each of the codecs under consideration. We can conclude that JPEG XL has lower maximum psychovisual difference according to this metric. The average value for JPEG XL is 57-69% of the average for other codecs.
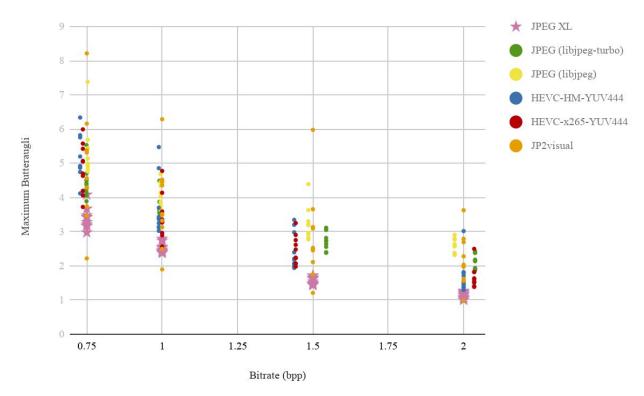
Fig. 9. Scatterplot of bitrate versus maximum Butteraugli error for each image and codec. Lower is better, 1.0 means indistinguishable from the original at a viewing distance of 1000 pixels. Note that metric values ≥ 6 indicate poor quality, and may be unordered with respect to other values ≥ 6.

Figure 10 has a similar layout but shows the p-norm instead of the maximum. Here, the average value for JPEG XL is 73-83% of the average value for HEVC-{HM|x265}-YUV444 and JP2visual.
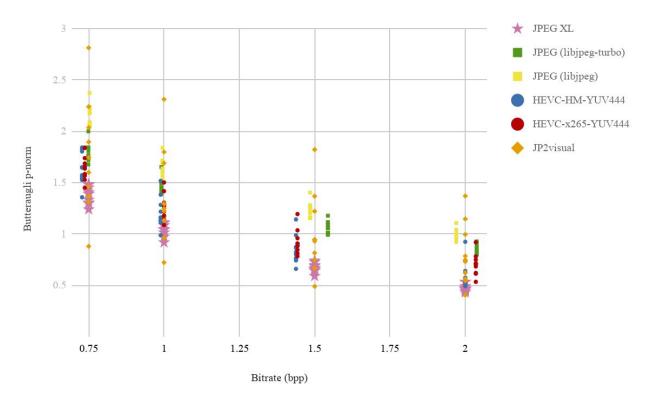
Fig. 10. Scatterplot of bitrate versus Butteraugli p-norm for each image and codec. Lower is better.

6.3 MS-SSIM

Multiscale SSIM is a measure of "distortion" in the sense of L2 norm while taking into account structural and contrast information in images. It is rather widely recognized by experts in imaging sciences that state of the art distortion metrics are not fully consistent with human perception. A recent paper formalized the finding that optimal rate/distortion performance does not imply perceptual quality. Under fairly mild assumptions (that perceptual quality could be decided with real/fake judgments by users, and expressed as divergence between two distributions), low distortion inevitably requires a higher bitrate and/or lower perceptual quality.[21]

The MS-SSIM results can be seen in Figure 11. JPEG XL behaves similarly to HEVC-HM-YUV444 according to this distortion metric, though at slightly higher bitrates. JP2visual often has higher MS-SSIM. However, note the limitations of this L2-like distortion metric. Low errors in large but unimportant areas such as backgrounds can outbalance severe artifacts elsewhere in the image. Hence we advocate metrics that report the *maximum* error, so that users can be confident that images are not unacceptably degraded anywhere. Recall that on average, the maximum Butteraugli distance of JP2visual is 1.5 times that of JPEG XL.
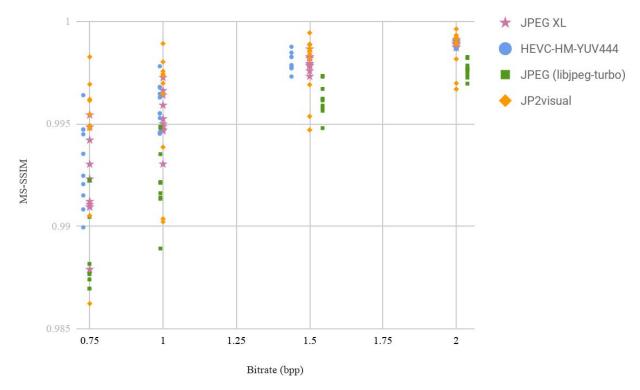
Fig. 11. Scatterplot of bitrate versus MS-SSIM result for each image and codec. Higher is better.

7. SOFTWARE ENCODE AND DECODE SPEED

7.1 Dataset and software

The input images are the set of 10 from Section 6.1 plus a 45.4 megapixel image (NC013_Evening_sRGB_8bit image). The latter is useful for evaluating scalability because it is large enough to ensure plentiful parallelism.

For JPEG XL, we used the open-source reference software (https://gitlab.com/wg1/jpeg-xl) at revision 0709f3 (February 19th, 2020). Butteraugli metrics are computed using the butteraugli_main (or equivalently, benchmark_xl) tools within the same repository.

For HEVC-HM-YUV444, we used version 16.20+SCM-8.8 of the reference encoder with no parallelism and no chroma subsampling (Y′ CbCr 4:4:4).

For HEVC-x265-YUV444, we used x265 v3.1.1 with no parallelism unless otherwise noted and no chroma subsampling, with the default speed setting of "medium". Note that x265 is only an encoder. We used the decoder in FFmpeg at revision 2d58fa (February 5th, 2020), which is not parallelized, even when the input would allow it.

For JPEG, we used libjpeg v1.56 and libjpeg-turbo v1.5.2.

7.2 Measurements on Intel

The test machine had 2 × Intel(R) Xeon(R) CPU E5-2690 v4 @ 2.60GHz (2×14 = 28 cores, 56 threads), with 128GB of DDR4 memory. The OS was Linux and the CPU governor was "performance".

The speeds were obtained by running each encoding or decoding command eight times, taking the geometric mean of elapsed time, and computing the speeds based on that. During each encoding or decoding task, the encoder or decoder was the only active process on the test machine.

7.2.1 JPEG XL scalability

For the NC013_Evening_sRGB_8bit image, the encoder was used at speed setting "cheetah", and at quality setting "d0.98" to match the quality (as measured by the Butteraugli metric) of libjpeg-turbo at quality 93. With these parameters, JPEG XL yielded a 4 647 021-byte file, which translates to 0.818 bits per pixel (bpp). Parallelization has no impact on the encoded file, nor on the decoded image. The encode and decode throughputs for various numbers of threads are shown in Table 12 and Figure 13. The command line used was the following, with $num_threads set to the number of threads to use:

```
tools/benchmark_xl --input .../NC013_Evening_sRGB_8bit.png \
    --codec jxl:d0.98:cheetah --encode_reps=8 --decode_reps=8 \
    --inner_threads=$num_threads
```

Tab. 12. Encode and decode throughput in megapixels/second for various thread counts on the NC013_Evening_sRGB_8bit image.

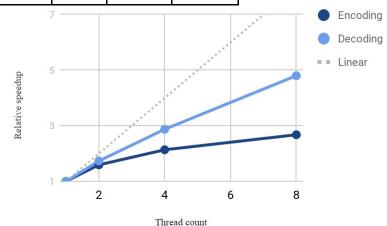| Threads	1	2	4	8
Encode MP/s	23.29	37.11	49.75	62.37
Decode MP/s	46.08	79.86	132.42	221.28

Fig 13. Scalability of JPEG XL on a single 45-megapixel image, relative to serial encoding and serial decoding.

For the 10 small images from Section 6.1, the speeds measured are listed in Table 14.

Tab. 14. encode and decode throughput in megapixels/second for various thread counts on the images from Section 6.1.

Threads	1	2	4	8
Encode MP/s	7.444	11.531	13.913	17.020
Decode MP/s	34.400	50.374	73.444	93.334

Parallelism was done within each encoding or decoding process. The images were processed sequentially. Note that the speeds and scalability is less than in the larger image because these images are smaller and thus offer fewer opportunities for parallel processing.

7.2.2 JPEG XL adaptive reconstruction

"Adaptive reconstruction" is a JPEG XL feature that helps achieve more pleasant images at lower bitrates but does not have a significant effect at the quality setting used here. In this section, we study its impact on encoding and decoding speed. Figure 15 shows encoding and decoding throughputs with and without that feature for various thread counts.

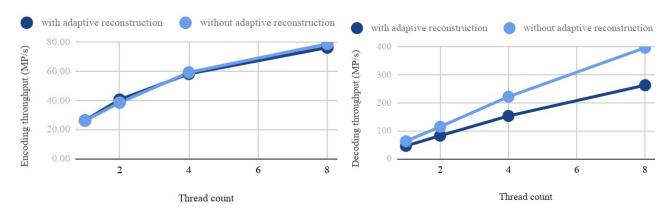The command lines used were as follows, with $num_threads set to the number of threads to use:

- ```
  tools/benchmark_xl --input .../NC013_Evening_sRGB_8bit.png \
      --codec jxl:d0.98:cheetah --encode_reps=8 --decode_reps=8 \
      --inner_threads=$num_threads
  ```
- ```
  tools/benchmark_xl --input .../NC013_Evening_sRGB_8bit.png \
      --codec jxl:d0.98:cheetah --encode_reps=8 --decode_reps=8 \
      --adaptive_reconstruction=0 --inner_threads=$num_threads
  ```


Fig. 15. Encoding (left) and decoding (right) throughput with and without adaptive reconstruction filter.

7.2.3 JPEG XL speed settings

To show the impact of the speed setting on density for a given target quality, we ran JPEG XL at speed settings "squirrel" and "kitten", targeting the same quality (as measured by the Butteraugli metric) reached by "cheetah". The command line used was the following, with $num_threads set to the number of threads to use per encoding or decoding task:

```
tools/benchmark_xl --input .../NC013_Evening_sRGB_8bit.png --codec \
    jxl:d0.98:cheetah,jxl:d0.96:squirrel,jxl:d1.01:kitten --encode_reps=8 \
    --decode_reps=8 --num_threads=1 --inner_threads=$num_threads
```

Table 16 reports the resulting bitrates and Figure 17 shows the encode and decode throughputs.

Tab. 16. Bitrates reached by various encoder speed settings on the NC013_Evening_sRGB_8bit image.

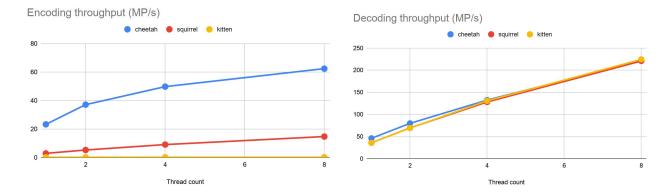| speed setting	bitrate (bpp)
cheetah	0.818
squirrel (default)	0.780
kitten	0.738

Fig. 17. Encoding and decoding throughput for various speed settings and thread counts
on the NC013_Evening_sRGB_8bit image.

7.2.4 Comparison with other codecs

For JPEG XL, we report results for 4 threads because quad-core systems are commonly available. Note that typical implementations of JPEG, HEVC and JPEG XL differ in terms of parallelization. libjpeg and libjpeg-turbo are not parallelized because this would require reset markers, which are not always present in JPEG bitstreams. HEVC can be split into independent parts at the risk of cross-border discontinuities. By contrast, JPEG XL includes tools for avoiding or reducing artifacts across part boundaries.

Table 18 shows the throughputs measured for each codec on the NC013_Evening_sRGB_8bit image. HEVC-HM-YUV444 is invoked with QP = 13 to match the JPEG XL bitrate (0.818 bpp), and HEVC-x265-YU444 with QP = 18, yielding a bitrate of 0.790 bpp. For JPEG, libjpeg-turbo with quality setting 93 yields a 9 565 605-byte file (1.684 bpp) that matches JPEG XL d0.98 in cheetah mode in terms of visual quality as assessed by the Butteraugli distance. Note that the JPEG XL encoder is 3 orders of magnitude faster than HEVC-HM-YUV444, and 13 times as fast as HEVC-x265, which we believe is more commonly used in practice than HM.

Tab. 18. Encode and decode throughputs in megapixels/second for the NC013_Evening_sRGB_8bit image.

Codec	Encode	Decode
JPEG XL cheetah (N=4)	49.753	132.424
JPEG (libjpeg)	9.013	11.133
JPEG (libjpeg-turbo)	48.811	107.981
HEVC-HM-YUV444	0.014	5.257
HEVC-x265-YUV444	1.031	14.037
HEVC-x265-YUV444 (N = 4)	3.691	14.100
HEVC-x265-YUV444 (N = 8)	6.345	13.471

Table 19 shows the throughputs measured for a subset of the codecs on the 10 images from Section 6.1. HEVC-HM-YUV444 is invoked with QP = 18 to match the JPEG XL overall bitrate.

Tab 19. Encode and decode throughputs in megapixels/second on the 10 images from Section 6.1.

Codec	Encode	Decode
JPEG XL cheetah (N=4)	13.913	73.444
JPEG (libjpeg)	8.502	10.360
JPEG (libjpeg-turbo)	38.456	94.311
HEVC-HM-YUV444	0.019	4.488

7.3 Measurements on ARM

The test machine is a Hikey970 with 4 Cortex-A73 cores at 2.36 GHz + 4 Cortex-A53 cores at 1.8 GHz in a "big.LITTLE" arrangement, and 6GB of LPDDR4X RAM. Table 20 shows the throughputs measured with JPEG XL at distance 0.98 and speed setting "cheetah" on the NC013_Evening_sRGB_8bit image.

Tab. 20. JPEG XL encode and decode throughputs in megapixels/second on NC013_Evening_sRGB_8bit depending on the set of CPU cores used. Left/right halves are with adaptive reconstruction enabled/disabled.

	4 little	4 big	1 little	1 big		4 little	4 big	1 little	1 big
Encode	5.91	19.25	1.83	6.56		5.97	19.49	1.84	6.60
Decode	24.75	47.75	6.17	10.87		36.02	76.59	9.17	18.38

8. CONCLUSION

This paper analyzed the performance of the novel JPEG XL codec in terms of perceptual visual quality and encoding/decoding speeds, at bitrates relevant for typical Internet usage (see data in Section 5.3). We also compared it with a set of popular modern image compression methods.

At bitrates near 1, both JPEG XL and HEVC are generally indistinguishable from the original according to the DCR-HR subjective evaluation methodology used here. Although the 95% confidence intervals often overlap, HEVC and JPEG XL codecs are significantly better (non-overlapping intervals) than JPEG 2000 at two rate points/images. We observe that HEVC produces more visually pleasing images at lower bitrates when distortions due to compression become clearly visible, whereas JPEG XL better maintains facial details and texture at typical bitrates corresponding to good quality.

To better understand the image quality at higher bitrates than DCR-HR can evaluate, we turn to objective quality metrics. JPEG XL reaches significantly lower maximum and p-norm Butteraugli distances than HEVC-{HM|x265}-YUV444 and JPEG 2000 at all bitrates evaluated: 0.75, 1, 1.5, 2. Note that JPEG XL as used here does not optimize for the Butteraugli metric. Insofar as Butteraugli characterizes perceptual quality, we conclude that JPEG XL reaches higher perceptual quality compared to the other codecs considered, at bitrates in common use on the Internet.

Modern codecs tend to have lower encoding and decoding throughput than JPEG. However, JPEG XL with 4 threads outperforms both libjpeg and libjpeg-turbo in encode and decode. Because HEVC-HM is not considered optimized for practical deployment, we also compared to the default setting of HEVC-x265. The default JPEG XL "squirrel" is 3 times

as fast as HEVC-x265, and JPEG XL "cheetah" is 13 times as fast (all using 4 threads). The JPEG XL decoder (again with 4 threads) is 9 times as fast as FFmpeg, which is not parallelized.

This paper reported high software encode/decode throughput and leading compression versus perceptual quality trade-offs, at a wide range of bitrates suitable for medium-quality up to visually (or even mathematically) lossless encoding of still images. JPEG XL offers key features not supported by contemporary (video-derived) codecs such as flexible progressiveness for responsive images, and a transition path (lossless transcoding) for the trillions of existing JPEG image files. We believe the performance and features of JPEG XL combine to make it the most favorable solution for internet images and high-quality photos.

ACKNOWLEDGEMENTS

REFERENCES

[1] Wallace, G. K., "The JPEG still picture compression standard," IEEE Trans. Consum. Electron. **38**(1), xviii–xxxiv (1992).

[2] Lee, D. T., "JPEG 2000: Retrospective and New Developments," Proc. IEEE **93**(1), 32–41 (2005).

[3] Christopoulos, C., Skodras, A. and Ebrahimi, T., "The JPEG2000 still image coding system: an overview," IEEE Trans. Consum. Electron. **46**(4), 1103–1127 (2000).

[4] Skodras, A., Christopoulos, C. and Ebrahimi, T., "The JPEG 2000 still image compression standard," IEEE Signal Process. Mag. **18**(5), 36–58 (2001).

[5] De Simone, F., Ouaret, M., Dufaux, F., Tescher, A. G. and Ebrahimi, T., "A comparative study of JPEG2000, AVC/H. 264, and HD photo," Opt. Eng. Appl., 669602–669602, International Society for Optics and Photonics (2007).

[6] De Simone, F., Ticca, D., Dufaux, F., Ansorge, M. and Ebrahimi, T., "A comparative study of color image compression standards using perceptually driven quality metrics," Opt. Eng. Appl., 70730Z–70730Z, International Society for Optics and Photonics (2008).

[7] De Simone, F., Goldmann, L., Baroncini, V. and Ebrahimi, T., "Subjective evaluation of JPEG XR image compression," Proc SPIE **7443**, 74430L (2009).

[8] Santa-Cruz, D., Grosbois, R. and Ebrahimi, T., "JPEG 2000 performance evaluation and assessment," Signal Process. Image Commun. **17**(1), 113–130 (2002).

[9] Ouaret, M., Dufaux, F. and Ebrahimi, T., "On comparing JPEG2000 and Intraframe AVC," SPIE Opt. Photonics, 63120U–63120U, International Society for Optics and Photonics (2006).

[10] Sullivan, G. J., Ohm, J.-R., Han, W.-J. and Wiegand, T., "Overview of the High Efficiency Video Coding (HEVC) Standard," IEEE Trans. Circuits Syst. Video Technol. **22**(12), 1649–1668 (2012).

[11] Lainema, J., Bossen, F., Han, W.-J., Min, J. and Ugur, K., "Intra Coding of the HEVC Standard," IEEE Trans. Circuits Syst. Video Technol. **22**(12), 1792–1801 (2012).

[12] Hanhart, P., Rerabek, M., Korshunov, P. and Ebrahimi, T., "Subjective evaluation of HEVC intra coding for still image compression," Seventh Int. Workshop Video Process. Qual. Metr. Consum. Electron.-VPQM 2013 (2013).

[13] Hanhart, P., Rerabek, M., De Simone, F. and Ebrahimi, T., "Subjective quality evaluation of the upcoming HEVC video compression standard," Proc SPIE **8499**, 84990V (2012).

[14] Zeng, W., Daly, S. and Lei, S., "An overview of the visual optimization tools in JPEG 2000," Signal Process. Image Commun. **17**(1), 85–104 (2002).

[15] "ITU-R BT.709 : Parameter values for the HDTV standards for production and international programme exchange."

[16] "ISO/IEC JTC 1/SC 29/WG1 M86083 JPEG XL Experiment Reproduction.", Sydney, Australia (2020).

[17] "ITU-R Rec. BT.500-13 Methodology for the subjective assessment of the quality of television pictures.", (2012).

[18] "ITU-R Rec. BT.2022-12 General viewing conditions for subjective assessment of quality of SDTV and HDTV television pictures on flat panel displays.", (2012).

[19] "ITU-REC-P.910: Subjective video quality assessment methods for multimedia applications.", (2008).

[20] Alakuijala, J., Asseldonk, R. van, Boukortt, S., Bruse, M., Comşa, I.-M., Firsching, M., Fischbacher, T., Kliuchnikov, E., Gomez, S., Obryk, R., Potempa, K., Rhatushnyak, A., Sneyers, J., Szabadka, Z., Vandervenne, L., Versari, L. and Wassenberg, J., "JPEG XL next-generation image compression architecture and coding tools," Appl. Digit. Image Process. XLII **11137**, 111370K, International Society for Optics and Photonics (2019).

[21] Blau, Y. and Michaeli, T., "Rethinking Lossy Compression: The Rate-Distortion-Perception Tradeoff," Int. Conf. Mach. Learn., 675–685 (2019).