

# Comprehensive Assessment of Image Compression Algorithms

Michela Testolina, Evgeniy Upenik, and Touradj Ebrahimi

Multimedia Signal Processing Group (MMSPG)  
Ecole Polytechnique Fédérale de Lausanne (EPFL)  
CH-1015 Lausanne, Switzerland  
Email: `firstname.lastname@epfl.ch`

## ABSTRACT

JPEG image coding standard has been a dominant format in a wide range of applications in soon three decades since it has been released as an international standard. The landscape of applications and services based on pictures has evolved since the original JPEG image coding algorithm was designed. As a result of progress in the state of the art image coding, several alternative image coding approaches have been proposed to replace the legacy JPEG with mixed results. The majority among them have not been able to displace the unique position that the legacy JPEG currently occupies as the ubiquitous image coding format of choice. Those that have been successful, have been so in specific niche applications. In this paper, we analyze why all attempts to replace JPEG have been limited so far, and discuss additional features other than compression efficiency that need to be present in any modern image coding algorithm to increase its chances of success. Doing so, we compare rate distortion performance of state of the art conventional and learning based image coding, in addition to other desired features in advanced multimedia applications such as transcoding. We conclude the paper by highlighting the remaining challenges that need to be overcome in order to design a successful future generation image coding algorithm.

**Keywords:** Image compression, lossy compression, deep learning, perceptual visual quality, JPEG standard

## 1. INTRODUCTION

During the last three decades, consumer and professional applications have increasingly relied on information in form of digital pictures, to the extent that it is hard to imagine any multimedia product or service that does not use pictures in one way or another. This has been largely possible thanks to the existence of the legacy JPEG format which was instrumental in creating an immense ecosystem of digital pictures that can be exchanged seamlessly between devices and within and across products and services, thanks to the ubiquity of the JPEG format. This ecosystem currently is estimated to produce over 10 billion shared JPEGs per day as a subset of an unknown number of digital pictures created in this same format, that could be up to two orders of magnitude large in their number.

It is difficult to show another example of digital technology next to JPEG compression which has remained so successful and dominant after three decades of existence, in the fast moving field of information and communication technologies. The reason behind longevity of the legacy JPEG format is not due to the lack of attempts to offer alternatives. In fact, the standardization committee behind the format, as well as other standard setting committees and industries have tried to unseat the dominance of the legacy JPEG by proposing alternatives, resulting in limited penetration, if any. In parallel, academic and research communities have been working hard to explore new and more advanced image coding approaches and strategies, often showing superior compression efficiency when compared to the good old legacy JPEG. However, among these more advanced compression technologies, so-far, none has been able to replace the legacy JPEG neither.

It is therefore natural to ask why, despite multiple attempts, legacy JPEG has not been replaced by better image compression formats, in the same way that in video compression, often older generation formats are to a large extent replaced by new and superior versions. The goal of this paper is to answer to this question.

The paper is organized as follows. After this introduction, Section 2 will provide an overview of image coding algorithms, both in terms of conventional as well as completely new paradigms for compression. Section 3 outlines the limitations of existing image coding algorithms and discusses why despite their superior compression efficiency, most did not succeed to reach their objective of replacing JPEG and why some had only limited success. Section 4 discusses some of the key technical challenges that should be overcome in order to increase chances of replacing the legacy JPEG by another ubiquitous and superior file format.

## 2. OVERVIEW OF IMAGE CODING ALGORITHMS

Usually, when people think of image compression, the JPEG standard is the first thing that comes to their mind. In fact, JPEG has been a major standard for image compression since 1992. Many more advanced coding algorithms have tried to replace it, but majority of us are still compressing images in the good old legacy JPEG format.

This section aims at summarizing the main coding algorithms for image compression. In particular, the JPEG and the subsequent conventional coding algorithms are presented. Furthermore, the new trend in image compression, namely learning-based image coding, algorithms are introduced.

### 2.1 JPEG and Successors

Standardized in 1992, the JPEG<sup>1</sup> (ISO/IEC 10928-1) is historically the first format for image compression. Its underlying algorithm is simple: it takes as input an image, divides it into 8x8 blocks, computes the DCT transform and quantizes the resulting coefficients with a pre-defined quantization table. After applying a zig zag scan, an entropy coding is applied to create a compressed bitstream. The JPEG proposed two different solutions for the latter: the first is IPR free and uses Huffman entropy coding; the second uses arithmetic coding but at the time it was introduced, it required a licensing fee. Despite the fact that the arithmetic coding mode is 10% more efficient, the popular implementation of JPEG uses the Huffman entropy coding, because it is free of charge.

Since the introduction of JPEG, many more advanced image compression algorithms have been proposed. The main conventional methods are:

- VTC:<sup>2</sup> proposed in 1999 by MPEG-4, VTC is the first image compression algorithm that attempted to replace JPEG. It aims at dealing with hybrid synthetic and natural contents (SNHC) and to support network applications. It proposed to use a wavelet transform as a replacement of DCT. The main problem of this approach, as explained by Norcen et al.,<sup>3</sup> is in its computational complexity, because it has a 5-12 times higher runtime requirements when compared to JPEG with marginal improvements in compression efficiency.
- JPEG 2000:<sup>4</sup> this standard was finalized in year 2000 by the Joint Photographic Experts Group and targets a large variety of applications. It introduces multiple features, mainly a performance improvement especially at low bit rates, region-of-interest coding and a better error resiliency. It also proposes a block(tiles)-based approach using DWT instead of DCT, generating more visually pleasant results. Regarding the entropy coding, it uses an arithmetic coding by default. The main drawback of the JPEG 2000 is in its computational complexity when compared to JPEG.
- HD Photo:<sup>5</sup> this algorithm was proposed in 2006 by Microsoft. It aims at addressing the demands of professional digital photography by supporting High Dynamic Range (HDR) as well as Standard Dynamic Range (SDR) in images. It is also compatible with multiple color spaces and supports transparency. One of the main novelties of HD Photo is in its use of LBT instead of DCT and a more advanced entropy coding.
- JPEG XR<sup>6</sup> (ISO/IEC 29199-2): was finalized in 2009, and it is very similar to HD Photo as it was based on the same technology and maintains interoperability with the latter.

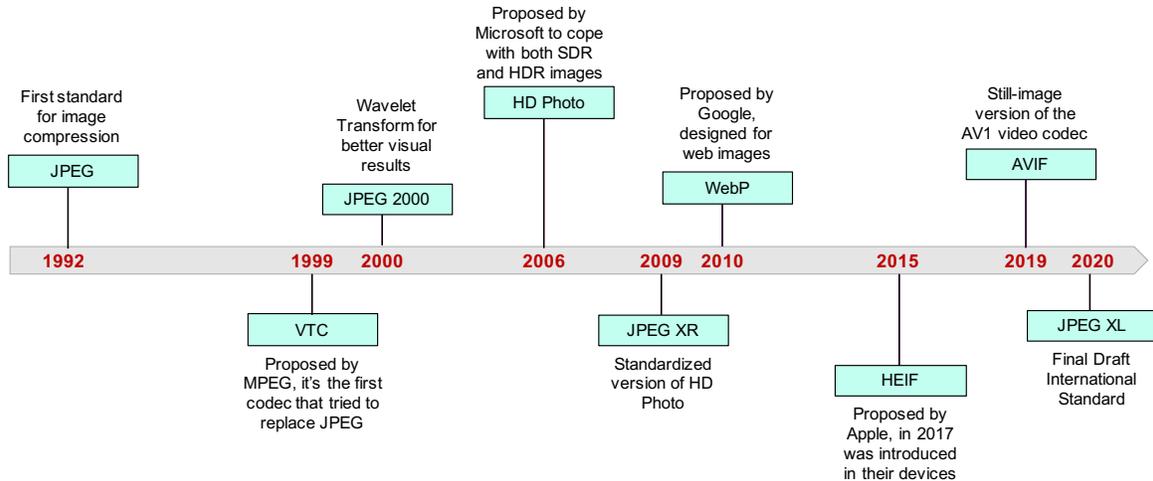


Figure 1: Timeline of conventional image coding algorithms with superior compression efficiency, in the last three decades

- WebP:<sup>7</sup> is a codec proposed in 2010 by Google, designed to target image compression for the world wide web applications. It supports new features such as animations. It was developed as the intra-frame coding profile of the video coding VP8 also proposed by Google. Differently to the previous coding algorithms, it exploits the correlation between neighbouring blocks using a block prediction scheme that aims at estimating the current block using the 4 previously-encoded neighbouring blocks. Then, the difference is encoded in a similar way to JPEG.
- HEIF:<sup>8</sup> was developed by the Moving Picture Expert Group (MPEG) and finalized in 2015. It is the still image compression profile of the video compression standard HEVC.<sup>9</sup> This standard introduces features such as compression of image sequences as in burst-photos. The standard is based on variable sized DCT and an intra-frame block-based prediction. Since 2017 the HEIF format has been included in Apple ecosystem, offering the possibility to take pictures and store them directly using this format.
- AVIF (AV1 Image File Format):<sup>10</sup> is the most recent image coding format, finalized in 2019 by an industry consortium called AOM. It is the intra-frame profile of the video compression standard AV1 also proposed by AOM. AVIF has been designed for web applications and aims at having the highest possible compression efficiency, especially at the low bit rates, to make the download of images from the web as quick as possible.

Figure 1 shows the timeline of the main conventional image compression algorithms with superior compression efficiency, introduced in the last three decades.

## 2.2 Learning-Based Image Compression

In the last few years, thanks to improvements in hardware technologies and new insights in machine learning, a number of image compression algorithms based on deep-learning have been proposed. These methods, in fact, shows promising performance in terms of compression efficiency and perceived visual quality.<sup>11</sup> Generally, a learning-based image coding compress the data using an autoencoder architecture, which is a type of neural network able to reduce the dimensionality of data in an unsupervised manner. It is usually composed of an encoder which transforms the data into a lower-dimensional latent representation and a decoder that reconstructs the source image from the latter. This approach is totally different from those used in conventional coding, and explores the problem from an entirely different perspective.

One of the first examples of image compression using autoencoders was presented in 2006 by Hinton et al.,<sup>12</sup> showing a non-linear autoencoder solution with multiple hidden layer and a strategy to effectively initialize its parameters.

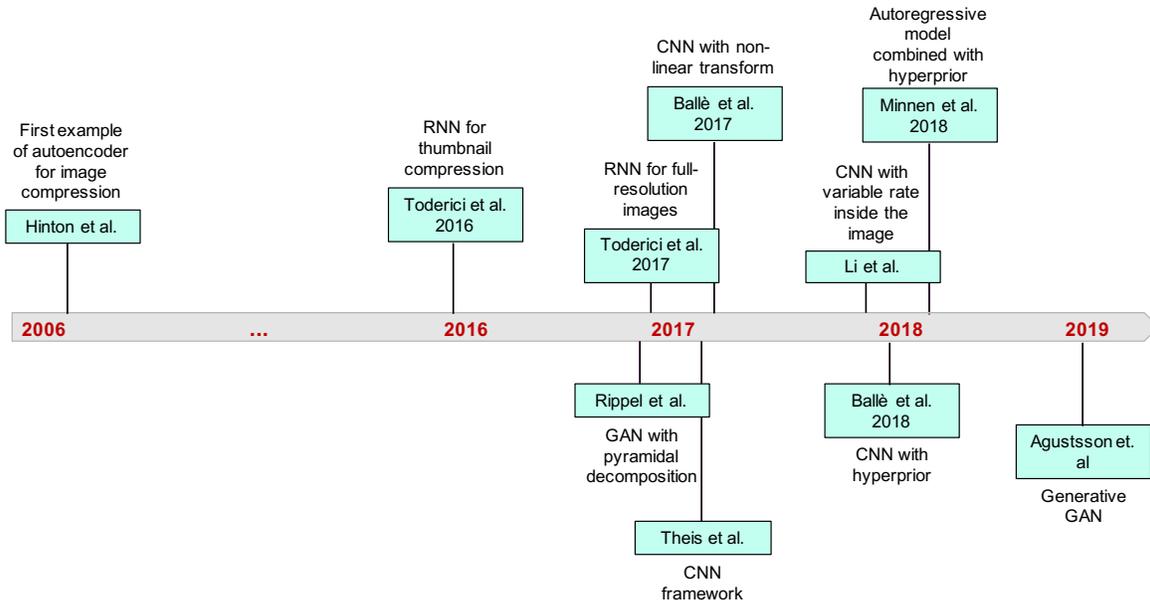


Figure 2: Timeline of learning-based coding for image compression

We can cluster learning-based image coding algorithms as a function of the neural network they rely on, as follows:

- Methods based on recurrent neural networks (RNN): the first example of RNN architecture for image compression was proposed by Toderici et al.<sup>13</sup> for thumbnail image compression, and extended by the same authors to full resolution images.<sup>14</sup> In their first work, the authors compared different types of RNN, showing promising results for convolutional/deconvolutional long short-term memory (LSTM) architecture. Based on this observation, in their second work the authors proposed a more advanced architecture based on an RNN encoder and decoder, a binarization technique and a novel entropy coder.
- Methods based on Generative Adversarial Networks (GANs): Rippel et al.<sup>15</sup> presented a GAN framework for image compression based on pyramidal decomposition. Consequently, Agustsson et al.<sup>16</sup> proposed a GAN framework for extremely low bit rates including a decoder/generator able to synthesize some details of the image, showing an improvement in the compression efficiency and in the visual quality of the images.
- Methods based on convolutional neural networks (CNN): Theis et al.<sup>17</sup> proposed a convolutional architecture that uses a smooth approximation and a novel entropy coding scheme. Consequently, Li et al.<sup>18</sup> presented a convolutional architecture for image compression with variable bit rate in different parts of an image based on a content-based importance map. A different approach was proposed by Ballè et al.<sup>19</sup> using nonlinear transforms (GDNs) in cascade to linear convolutions. This method was then extended using an hyperprior<sup>20</sup> to exploit the spatial dependencies in the latent representation and by Minnen et al.<sup>21</sup> introducing an autoregressive model combined with the hyperprior.

The timeline of the development of the learning-based coding for image compression is presented in figure 2.

Even if these methods have demonstrated to possess the potential of achieving good performance, their development is just at the beginning and they need further improvements before reaching maturity. Recently, the Joint Photographic Experts Group (JPEG) committee introduced an exploration for Learning-based Image Coding (JPEG AI) which aims at studying and developing Learning-Based image compression solutions in view of standardization. This group recently organized a Call for Evidence to demonstrate the potential of deep-learning approaches for image compression.

### 2.3 Rate-Distortion Comparison Tests

To assess the performance of codecs presented in the previous subsections, we visually compared their performance using test images from the dataset in Appendix A, after compressing them with 5 different conventional and 3 different learning-based image coding algorithms. Further information about the specific implementations used in each case are shown in table 1.

Table 1: List of the implementations used for rate-distortion comparison tests.

Codec	Implementation	Link	Remarks
JPEG	JPEG XT reference software v1.5	<a href="http://jpeg.org/jpegxt/software.html">http://jpeg.org/jpegxt/software.html</a>	
JPEG 2000	Kakadu, v8.0.5 for MacOS	<a href="https://kakadusoftware.com/downloads/">https://kakadusoftware.com/downloads/</a>	Visually Optimized Method
WebP	libwebp v.1.1.0 for MacOS	<a href="https://developers.google.com/speed/webp/download">https://developers.google.com/speed/webp/download</a>	
JPEG XR	JPEG XR reference software 1.41 (May 2012)	<a href="https://jpeg.org/jpegxr/software.html">https://jpeg.org/jpegxr/software.html</a>	
HEIF	Apple implementation in MacOS Mojave 10.14.6	<a href="https://jpeg.org/jpegxr/software.html">https://jpeg.org/jpegxr/software.html</a>	
Toderici et al.	Implementation available in Github, commit 12112ec	<a href="https://github.com/zhang-yi-chi/residual-rnn">https://github.com/zhang-yi-chi/residual-rnn</a>	Implementation of the paper <sup>14</sup>
Ballè et al.	Implementation available in Github, commit 9cbc850	<a href="https://github.com/tensorflow/compression/tree/master/models">https://github.com/tensorflow/compression/tree/master/models</a>	Implementation of the paper <sup>20</sup>
Minnen et al.	Implementation available in Github, commit 9cbc850	<a href="https://github.com/tensorflow/compression/tree/master/models">https://github.com/tensorflow/compression/tree/master/models</a>	Implementation of the paper <sup>21</sup>

We present the visual results in figure 3 and 4, showing a cropped section of one of the images to better appreciate the differences.

Observing the results in figure 3, it is possible to notice that the legacy JPEG introduces blocking artifacts in images, especially at target bit rates lower than 1.5bpp. Moreover for bit rates lower than 0.4bpp we observe a loss in the chrominance. All other codecs present more visually pleasant results. Among them, HEIF introduces the least artifacts, even at low bit rates. This example demonstrates that the legacy JPEG has the least efficient rate-distortion performance; but it is still the most popular codec used for image compression.

Figure 4 shows the results of the leaning-based codecs compared to the JPEG. These results reveal a big improvement in the performance of learning-based methods over the last few years. In particular the RNN proposed by Toderici et al.<sup>14</sup> in 2017 was one of the first examples of an end-to-end image compression and presents similar performance to the legacy JPEG; as a follow up, the architecture introduced by Ballè et al. in the early 2018<sup>20</sup> shows a big improvement in the performance. This last method was again further improved in late 2018 by Minnen et al.<sup>21</sup> These results also indicate that learning-based approach to image compression is just at an early stage and there is no evidence that such methods have reached maturity, as their performance are still improving at a rapid pace. Observing the improvement in the performance as showed in this example, it is possible to conclude that the learning-based approach for image compression are promising as future candidate to replace legacy JPEG.

### 3. LIMITATIONS OF CURRENT IMAGE COMPRESSION ALGORITHMS

In the previous section, we summarized the codecs for image compression. We showed with illustrative visual comparison examples that among all popular codecs, legacy JPEG presents the worst compression efficiency. So

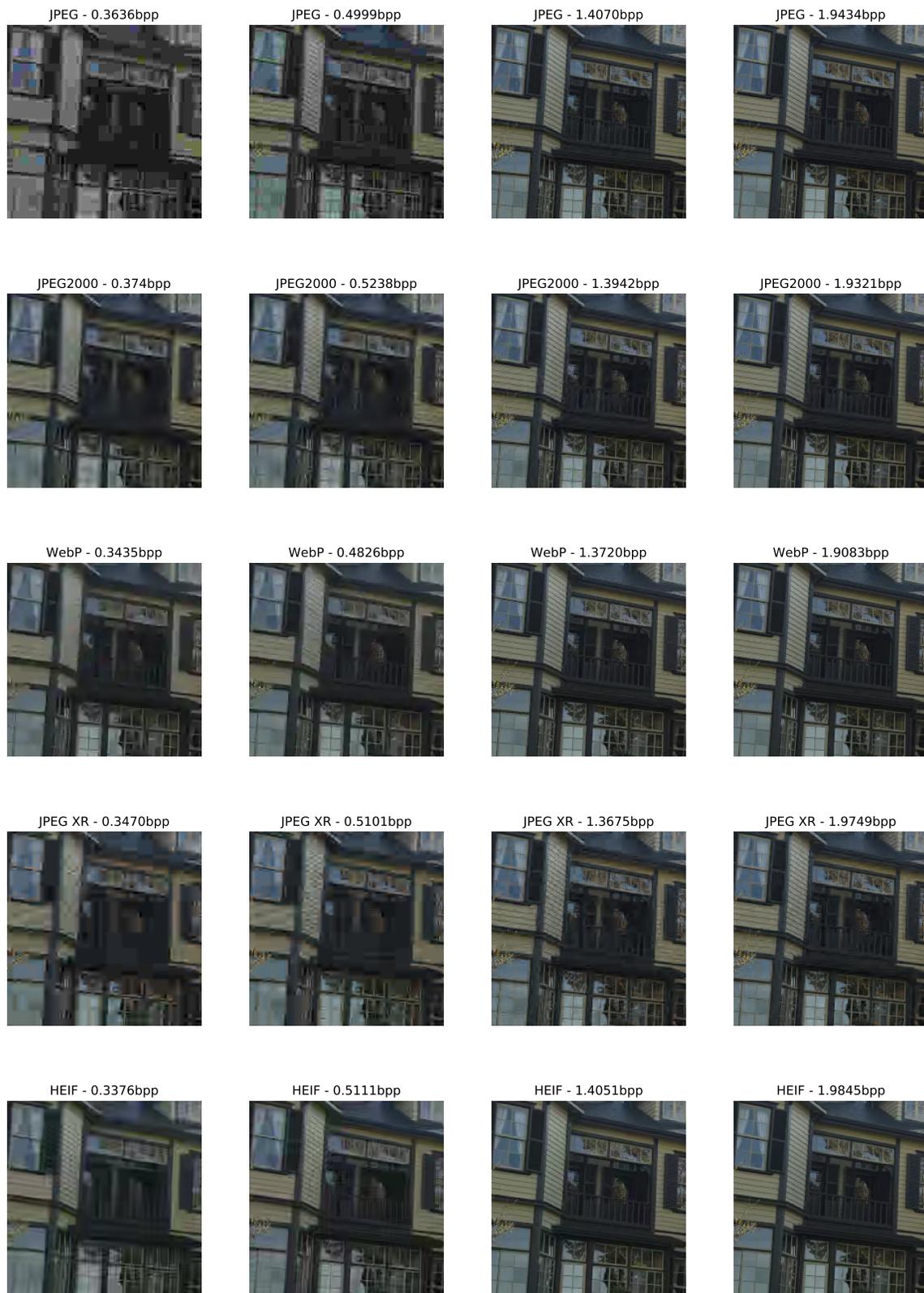


Figure 3: Visual comparison of different conventional codecs at different target bit rates.



Figure 4: Visual comparison of different learning-based codecs and the JPEG at different bit rates.

why is the legacy JPEG still the most popular codec for image compression? The reason could be behind some limitations in the alternative codecs. This section aims at better understanding such limitations.

### 3.1 Unified Image and Video Compression

The first limitation that one observes in alternative image compression algorithms is that many are driven from designs primarily targeted at video compression. For example, HEIF that has now been widely adopted in Apple devices, is the still-picture profile of High Efficiency Video Compression Standard (HEVC).

However, image compression often differs from video compression from many perspectives. Applications relying on image and video compression, in fact, usually have different requirements: while video compression often aims at optimizing the bit rate for the best possible quality, because in its typical use cases the bandwidth is limited (e.g. streaming, video conferencing, etc.), image compression instead, often aims at maximizing the quality as the best possible bit rate, as in most imaging applications, there is little appetite to exchange quality in favor of saving storage space which is more often unused than saturated.

Moreover, image and video compression algorithms exploit different properties of the Human Visual System (HVS): while in video compression the spatio-temporal sensitivity is mostly exploited, in image coding the spatial, chromatic and contrast sensitivity are taken into account in their design (e.g. noise sensitivity in different spatial frequencies and masking).

Furthermore, the way pictures and video contents are consumed and exploited are different in essence because of the static versus dynamic nature of each type of content.

Considering the above observations, efficient image and video compression algorithms should not be designed by constraining that their underlying architectures should be the same. In fact, although wavelet-based compression has shown to result in efficient image coding, the same has not been demonstrated in video compression despite many decades of attempts.

### 3.2 Lossy Transcoding

We already introduced that the JPEG has been the dominant format in image compression since 1992. As a result, trillions of images have been compressed and stored using this format. For this reason, a new compression algorithm should not only be able to efficiently compress uncompressed images, but also capable of converting JPEG-compressed images into the new format without any loss in their quality. This process is called transcoding. Such a conversion is not trivial as legacy JPEG is a lossy compression algorithm and intrinsically contains compression noise.

For example, the transcoding from legacy JPEG to HEIF as implemented in Apple devices, converts an existing JPEG image into HEIF by guaranteeing a lower storage at a targeted value, for instance half the size of the JPEG version. Performance of the transcoding algorithm was tested in an experiment on the uncompressed images from the dataset in Appendix A and converted into JPEG with different quality factors and then transcoded from JPEG to HEIF. When transcoding into HEIF, the targeted bit rate was set at half the bit rate of the JPEG images, based on the observation that \* images with an equivalent visual quality in HEIF are half the size of their counterparts in JPEG format. More precisely, to convert from raw to JPEG format, the JPEG XT reference software v1.5 † was used at multiple Q factors (10, 20, 30, 40, 50, 60, 70, 80 and 90). To convert from JPEG to HEIF, the Apple implementation in MacOS Mojave 10.14.6 was used.

All images of the testset were compressed, each with the 10 targeted qualities. The visual results for the images "House" and "Sailboat" are shown in figure 6 and 7. For both images one can observe that the transcoding introduces noise in areas with high-frequency and a loss of details resulting in lower quality. The degradation in the quality is more evident in the image "House" than in "Sailboat", as the former contains more high-frequencies.

Objective quality metrics were also calculated between the original uncompressed and the JPEG-compressed, and between the original uncompressed and the HEIF-transcoded images at half the bit rate of the JPEG-compressed images. In particular, PSNR was selected as it is a popular metric for quality assessment, and

---

\*<https://developer.apple.com/videos/play/wwdc2017/503/>

†<http://jpeg.org/jpegxt/software.html>

MS-SSIM as it is believed that it correlates better with human perception of quality. Figure 5 shows the results for 6 of the test images. Both PSNR and MS-SSIM metrics show a degradation of the quality for the HEIF transcoded images (orange curve), for all the images in the test-set, except for the screenshot.

The visual results show that transcoding is still an unsolved problem. An efficient compression algorithm should be able to transcode the legacy JPEG-compressed images without degrading the quality of the source or introducing additional artifacts, and in fact ideally even improving the quality of the source if possible.

### 3.3 Numerically Lossless Transcoding

Another attempt to cope with the plethora of JPEG images distributed over global computer networks and stored on private consumer devices has been proposed by JPEG XL.<sup>22,23</sup> This emerging standard, which is undergoing the final steps of standardization, features numerically lossless transcoding of images that are already compressed with JPEG. JPEG XL can save up to 22%<sup>22</sup> in bit rate or in storage when numerically lossless transcoding is applied to images compressed with JPEG, while preserving exactly the same visual quality.

In order to quantitatively assess the performance of JPEG XL numerically lossless transcoding of JPEG compressed images, we used the data that was obtained during subjective visual quality evaluation of JPEG and JPEG XL codecs performed in the test laboratory of Multimedia Signal Processing Group at EPFL. The experiment was conducted using the same dataset of original uncompressed images as that referred to earlier in this paper. Two codecs were assessed: namely, JPEG and JPEG XL. The latter was used to compress original images as well as to transcode in a numerically lossless manner the images that had been already encoded with the former.

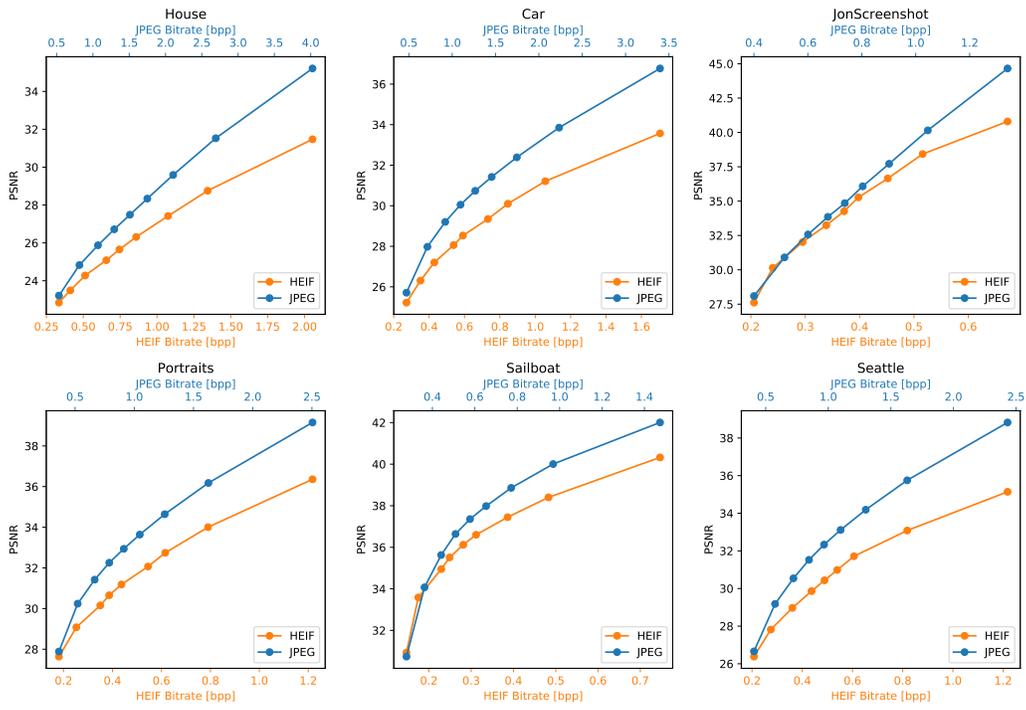
The subjective evaluation experiments were conducted in accordance with the ITU recommendations.<sup>24</sup> The monitor used in the experiments was EIZO CG301W. It has a resolution of 2560x1600 with a pixel density of 101 ppi and an 8-bit panel. The monitor was calibrated in compliance with the ITU recommendations according to the following profile: sRGB Gamut, D65 chromaticity of the background, 80 cd/m<sup>2</sup> 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. In order to evaluate the subjective visual quality, the degradation category rating (DCR-HR) method<sup>25</sup> with hidden reference and simultaneous presentation was used. 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 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 visual acuity examination and Ishihara test for color perception.

In Figure 8, one can find the results of the assessment of numerically lossless transcoding from JPEG to JPEG XL. It is important to point out that the transcoded images are bit-exactly identical to the reconstructed JPEG images. Thus, one can associate the subjective mean opinion scores obtained for JPEG with the results of numerically lossless JPEG XL transcoding. This way, in Figure 8, the orange plot (representing JPEG transcoded to JPEG XL) is a version of the blue plot (representing JPEG) that is shifted to the left. This shift shows the gain in the bit rate that is achieved while preserving exactly the same visual quality.

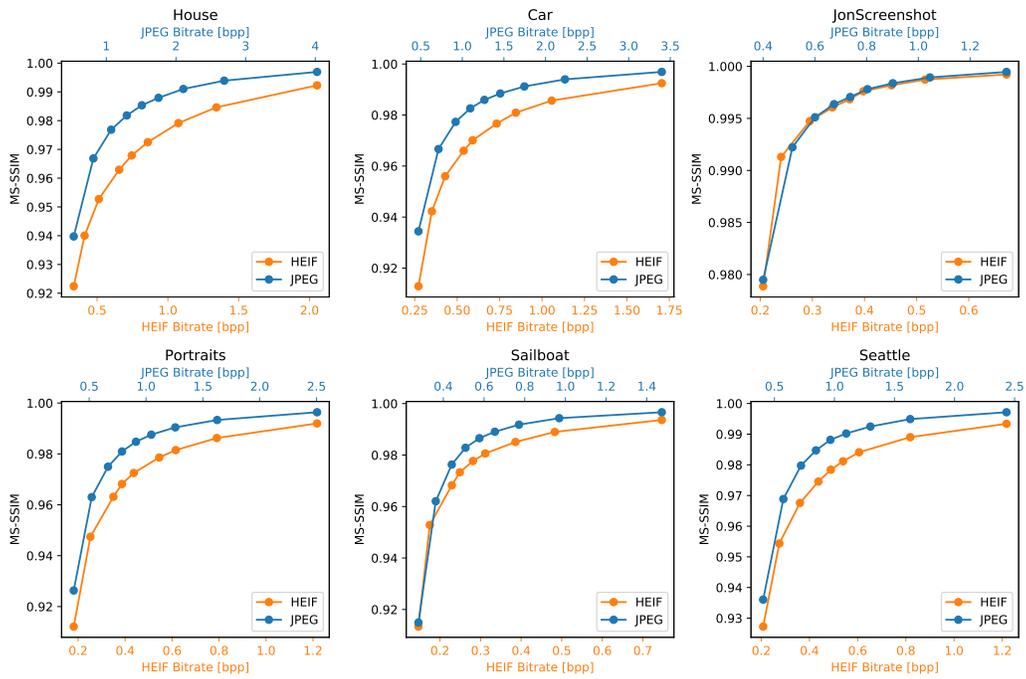
Numerically lossless transcoding implemented in JPEG XL standard provides a solution that helps coping with transition from legacy JPEG to JPEG XL. This approach can be used in the foreseeable future as a smooth transition before complete adoption of JPEG XL as a dominant standard replacing JPEG. But such a lossless transcoding is not the most optimal solution in terms of rate-distortion performance.

## 4. TOWARD THE NEXT GENERATION OF IMAGE CODING

In section 3, some limitations of existing codecs were discussed, focusing on the unified image and video compression and on issues regarding transcoding. In this section, we discuss open problems in image compression that needs to be solved to design efficient image compression beyond the state of the art.

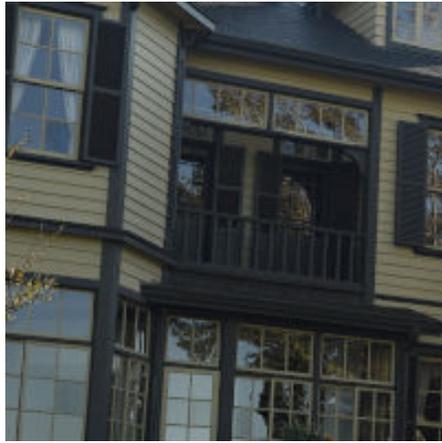


(a) PSNR



(b) MS-SSIM

Figure 5: PSNR and MS-SSIM metrics plotted against the bit rate.



(a) JPEG, 2.72bpp



(d) HEIF, 1.36bpp



(b) JPEG, 1.76bpp



(e) HEIF, 0.88bpp



(c) JPEG, 1.28bpp



(f) HEIF, 0.64bpp

Figure 6: Visual results of the transcoding for the image "House". The left column shows the artifacts in a cropped region generated by converting uncompressed to JPEG with 3 different quality factors. On the right, the same cropping was transcoded from JPEG to HEIF by preserving the same visual quality.



(a) JPEG, 0.96bpp



(d) HEIF, 0.48bpp



(b) JPEG, 0.64bpp



(e) HEIF, 0.32bpp



(c) JPEG, 0.48bpp



(f) HEIF, 0.24bpp

Figure 7: Visual results of the transcoding for the image "Sailboat". The left column shows the artifacts in a cropped region generated by converting uncompressed to JPEG with 3 different quality factors. On the right, the same cropped region was transcoded from JPEG to HEIF by preserving the same visual quality.

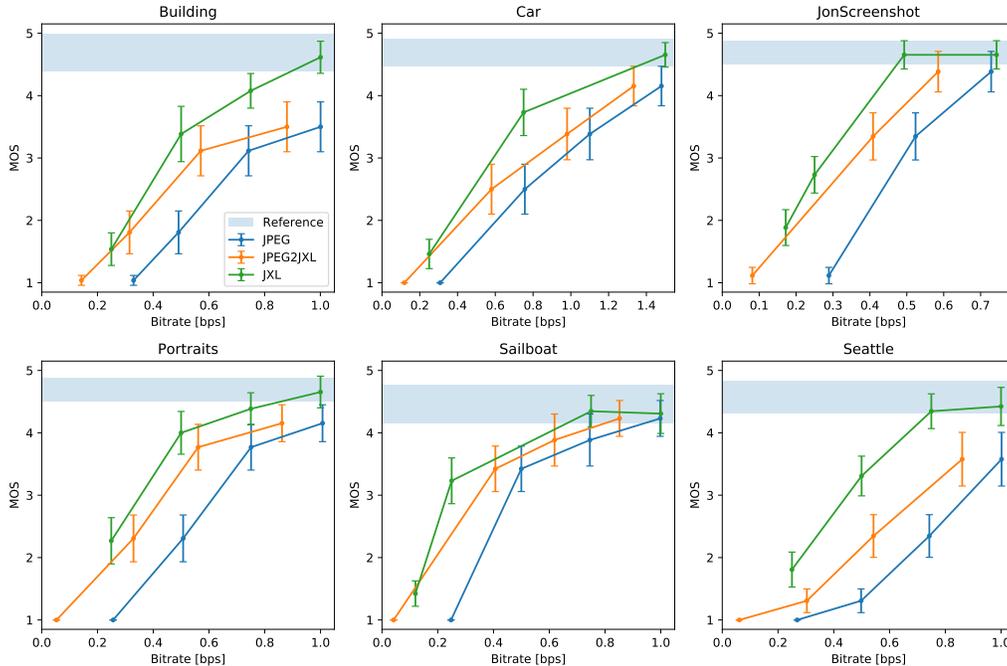


Figure 8: Mean opinion scores (MOS) and 95% confidence intervals (CI) with respect to stimulus bit rates for six test contents assessed subjectively. Blue line represents JPEG encoded stimuli. Orange line shows the data for numerically lossless transcoding of JPEG images to JPEG XL. Green line represents stimuli encoded with JPEG XL using original uncompressed test images as an input. The filled blue area represents 95% confidence interval for hidden reference, and may be considered as a transparent visual quality range.

#### 4.1 Reliable No-Reference Quality Metrics

The first open problem is the need for reliable no-reference quality metrics. No-reference quality metrics target at assessing the quality of distorted images without knowing the original uncompressed image. This quality assessment approach can be used in a more efficient transcoding process: a reliable no-reference metric can guide the transcoding process, by assessing the quality before and after transcoding, as shown in figure 9. However, the state of the art in no-reference quality assessment is still in its infancy and thus remains a challenge.

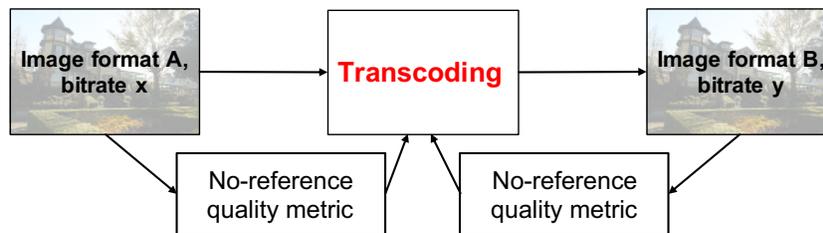


Figure 9: Visual representation of the transcoding process using no-reference quality metrics

#### 4.2 Transparent Visual Quality Metrics

In section 3.1 it was already observed that some of the current image codecs derive from video compression. It was also discussed that while video compression aims at primarily minimizing bit rate for the highest possible quality, image compression often optimizes bit rate for a targeted quality. Generally users prefer to have high

quality pictures because not only most represent special moments in life, but also scarcity of storage space is often less constraining than scarcity of communication bandwidth.

Taking into account such requirements, one can say that a good image compression algorithm should target at compressing images without introducing any visible distortions. A compressed image whose visual quality appears the same as the uncompressed is said to have a *transparent quality*. Any attempt in designing an algorithm that targets at generating compressed images with transparent visual quality should use a quality metric that is able to detect the transparent quality.

Transparent visual quality can be measured both subjectively and objectively. In the next subsections we discuss both and explain why they are still open problems.

#### 4.2.1 Subjective Transparent Visual Quality Methodologies

The first step in the design of a transparent visual quality metric is to be able to assess it subjectively. A early work in this area has been implemented in the framework of JPEG AIC Part 2 (ISO/IEC 29170-2),<sup>26</sup> by proposing a standardized subjective quality assessment methodology that can detect transparent visual quality. Despite this, the problem is still largely unexplored. A good methodology to assess the transparent visual quality of compressed images is fundamental to the design of an efficient objective transparent visual quality metric.

#### 4.2.2 Objective Transparent Visual Quality Metrics

Visual transparency can also be assessed objectively, through a transparent visual quality metric. Currently, no one has attempted to solve this problem. The existing objective quality metrics, in fact, aim at quantifying the differences between two images giving a continuous score that correlates to the perceived differences in quality. A transparent visual quality metric should instead propose a binary score that informs whether the differences between two images are perceivable by a human observer.

Designing a metric able to discriminate transparency in quality is not trivial and should consider different factors:

- content: the ability to discriminate transparency depends on the content of the image. For example, noise is usually more perceived in flat areas of images. So an image with a lot of high-frequencies is going to have a different transparency threshold than an images with a lot of low-frequency areas;
- display: the ability to discriminate transparency also depends on the screen. For example some compression artifacts might be visible in a full-HD monitor but not on a smartphone. Another important factor to consider is the distance between the observer and the display;
- context: some environmental factors could also influence the ability to perceive distortions. For example some image artifacts might not be visible in presence of a bright ambient light. Moreover, the ability to discriminate transparency is also different from person to person and depends on different factors like age, visual acuity, experience, etc.

Currently, none of the existing quality metrics target this specific challenge. The notable exception is in VMAF where a related idea is proposed,<sup>27,28</sup> through training with subjective data. It is thus possible to identify a threshold for transparency. However, VMAF doesn't propose to assess transparent visual quality directly.

## 5. CONCLUSIONS

Despite several attempts to replace the soon three-decade old JPEG legacy format, the latter remains the dominant choice in consumer imaging applications. One of the reasons for limited success of alternative image coding formats based on more advanced state of the art technologies, is that although they offer better compression efficiency, they lack other essential features that are important in order to enable a smooth transition of the dominant legacy JPEG format to a more recent, more efficient and more versatile image coding format. Two of these features are the ability to show superior compression efficiency in the operational points where the

quality of the image remains indistinguishable from its uncompressed version and a transcoding approach that can guarantee that the change of format and bit rate does not deteriorate the quality of the source image. The key to a successful image coding format is therefore in the design of reliable quality assessment metrics that can steer both compression parameters of a codec and transcoding it from another format (often legacy JPEG) to the new format. The latter when combined with machine learning based image compression approaches set the ground for a likely future where the legacy JPEG can be finally retired and replaced by another ubiquitous and dominant image coding format. In this paper, we did not address other technical and non-technical key influencing factors that are also essential in the success of any future more advanced image compression format, such as complexity, power consumption, adoption by important actors in the imaging industry and services as well as in popular applications, and last but not least licensing terms including existence of a base line with license fee free arrangement. A closer analysis of these other key factors are out of the scope of this paper but nevertheless also play an important role in the adoption of any image coding format in the future.

## ACKNOWLEDGMENTS

Authors would like to acknowledge contributions from the Swiss National Scientific Research project entitled "Advanced Visual Representation and Coding in Augmented and Virtual Reality" under grant number 200021\_178854 and Innosuisse project "Deep Fake Factory" under grant number 34270.1 IP-ICT.

## REFERENCES

- [1] Wallace, G. K., "The jpeg still picture compression standard," *IEEE transactions on consumer electronics* **38**(1), xviii–xxxiv (1992).
- [2] Moccagatta, L. and Chen, H., "Mpeg-4 visual texture coding: more than just compression," in [*1999 Digest of Technical Papers. International Conference on Consumer Electronics (Cat. No. 99CH36277)*], 302–303, IEEE (1999).
- [3] Norcen, R. and Uhl, A., "Performance issues in mpeg-4 vtc image coding," in [*Visual Communications and Image Processing 2003*], **5150**, 880–890, International Society for Optics and Photonics (2003).
- [4] Christopoulos, C., Skodras, A., and Ebrahimi, T., "The jpeg2000 still image coding system: an overview," *IEEE transactions on consumer electronics* **46**(4), 1103–1127 (2000).
- [5] Srinivasan, S., Tu, C., Regunathan, S. L., and Sullivan, G. J., "Hd photo: a new image coding technology for digital photography," in [*Applications of Digital Image Processing XXX*], **6696**, 66960A, International Society for Optics and Photonics (2007).
- [6] Dufaux, F., Sullivan, G. J., and Ebrahimi, T., "The jpeg xr image coding standard [standards in a nutshell]," *IEEE Signal Processing Magazine* **26**(6), 195–204 (2009).
- [7] Ginesu, G., Pintus, M., and Giusto, D. D., "Objective assessment of the webp image coding algorithm," *Signal Processing: Image Communication* **27**(8), 867–874 (2012).
- [8] Hannuksela, M. M., Lainema, J., and Vadakital, V. K. M., "The high efficiency image file format standard [standards in a nutshell]," *IEEE Signal Processing Magazine* **32**(4), 150–156 (2015).
- [9] ISO/IEC 23008-12:2017 Information technology — High efficiency coding and media delivery in heterogeneous environments — Part 12: Image File Format.
- [10] Chen, Y., Murherjee, D., Han, J., Grange, A., Xu, Y., Liu, Z., Parker, S., Chen, C., Su, H., Joshi, U., et al., "An overview of core coding tools in the av1 video codec," in [*2018 Picture Coding Symposium (PCS)*], 41–45, IEEE (2018).
- [11] Ascenso, J., Akyazi, P., Pereira, F., and Ebrahimi, T., "Learning-based image coding: early solutions reviewing and subjective quality evaluation," in [*Optics, Photonics and Digital Technologies for Imaging Applications VI*], **11353**, 113530S, International Society for Optics and Photonics (2020).
- [12] Hinton, G. E. and Salakhutdinov, R. R., "Reducing the dimensionality of data with neural networks," *science* **313**(5786), 504–507 (2006).
- [13] Toderici, G., O'Malley, S. M., Hwang, S. J., Vincent, D., Minnen, D., Baluja, S., Covell, M., and Sukthankar, R., "Variable rate image compression with recurrent neural networks," in [*International Conference on Learning Representations, ICLR 2016, San Juan, Puerto Rico*], (2016).

- [14] Toderici, G., Vincent, D., Johnston, N., Jin Hwang, S., Minnen, D., Shor, J., and Covell, M., “Full resolution image compression with recurrent neural networks,” in [*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*], 5306–5314 (2017).
- [15] Rippel, O. and Bourdev, L., “Real-time adaptive image compression,” in [*Proceedings of the 34th International Conference on Machine Learning, Sydney, Australia*], (2017).
- [16] Agustsson, E., Tschannen, M., Mentzer, F., Timofte, R., and Gool, L. V., “Generative adversarial networks for extreme learned image compression,” in [*Proceedings of the IEEE International Conference on Computer Vision*], 221–231 (2019).
- [17] Theis, L., Shi, W., Cunningham, A., and Huszár, F., “Lossy image compression with compressive autoencoders,” in [*International Conference on Learning Representations, ICLR 2017, Toulon, France*], (2017).
- [18] Li, M., Zuo, W., Gu, S., Zhao, D., and Zhang, D., “Learning convolutional networks for content-weighted image compression,” in [*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*], 3214–3223 (2018).
- [19] Ballé, J., Laparra, V., and Simoncelli, E. P., “End-to-end optimized image compression,” in [*International Conference on Learning Representations, ICLR 2017, Toulon, France*], (2017).
- [20] Ballé, J., Minnen, D., Singh, S., Hwang, S. J., and Johnston, N., “Variational image compression with a scale hyperprior,” in [*International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada*], (2018).
- [21] Minnen, D., Ballé, J., and Toderici, G. D., “Joint autoregressive and hierarchical priors for learned image compression,” 10771–10780 (2018).
- [22] Alakuijala, J., Asseldonk, R. v., Boukourt, S., Bruse, M., Comşa, I.-M., Firsching, M., Fischbacher, T., Kliuchnikov, E., Gomez, S., Obryk, R., Potempa, K., Rhatushnyak, A., Sneyers, J., Szabadka, Z., Vanderveen, L., Versari, L., and Wassenberg, J., “JPEG XL next-generation image compression architecture and coding tools,” in [*Applications of Digital Image Processing XLII*], **11137**, 111370K, International Society for Optics and Photonics (Sept. 2019).
- [23] Alakuijala, J., Boukourt, S., Ebrahimi, T., Kliuchnikov, E., Sneyers, J., Upenik, E., Vanderveen, L., Versari, L., and Wassenberg, J., “Benchmarking JPEG XL image compression,” in [*Optics, Photonics and Digital Technologies for Imaging Applications VI*], **11353**, 113530X, International Society for Optics and Photonics (Apr. 2020).
- [24] “ITU-R Rec. BT.500-13 Methodology for the subjective assessment of the quality of television pictures,” (2012).
- [25] “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).
- [26] ISO/IEC 29170-2:2015 Information technology — Advanced image coding and evaluation — Part 2: Evaluation procedure for nearly lossless coding.
- [27] Li, Z., Aaron, A., Katsavounidis, I., Anush, M., and Manohara, M., “Toward a practical perceptual video quality metric,” (2017).
- [28] Li, Z., Bampis, C., Novak, J., Aaron, A., Swanson, K., Moorthy, A., and Cock, J. D., “Vmaf: The journey continues,” (2018).

## APPENDIX A. DATASET OF IMAGES

Figure 10 shows seven original images selected for the evaluation experiments. All images have a bit-depth of 8 bits per channel, are in sRGB color space and represented in lossless PNG format. In Table 2, one can find the spatial resolution for each image in the dataset. The images were selected in order to represent different types of photographic scenes, such as landscapes, people, and objects. One non-photographic image was also included in the dataset.

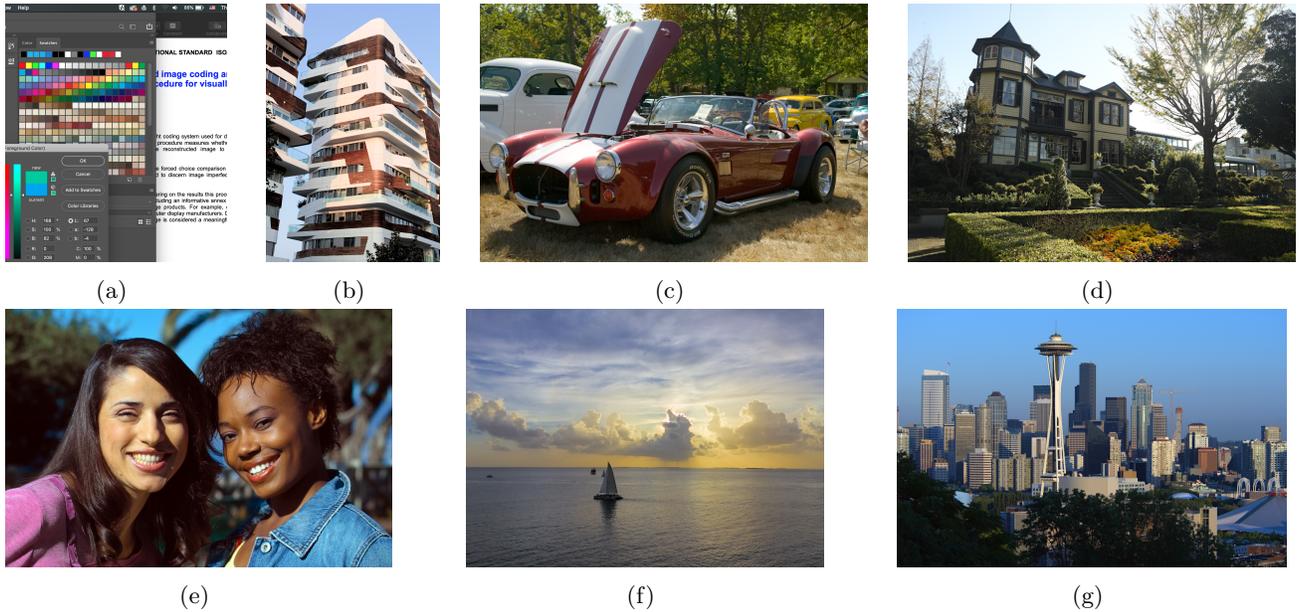


Figure 10: Contents selected for experiments. From left to right: "JonScreenshot", "Building", "Car", "House", "Portraits", "Sailboat", "Seattle".

Table 2: Resolution of selected contents

File	Resolution
nonphoto_JonScreenshot_1200x1400_8bit_sRGB.png	1200x1400
photo_Building_938x1400_8bit_sRGB.png	938x1400
photo_Car_1016x676_8bit_sRGB.png	1016x676
photo_House_1032x688_8bit_sRGB.png	1032x688
photo_Portraits_1134x756_8bit_sRGB.png	1134x756
photo_Sailboat_1200x866_8bit_sRGB.png	1200x866
photo_Seattle_1072x712_8bit_sRGB.png	1072x712