CEPAT Journal of Computer Engineering: Progress, Application and Technology
Vol. 5, No. 2, November 2026, pp. 206~215
ISSN: 2963-6728, DOI: https://doi.org/ 10.25124/cepat.v5i02.12430

Trade-off Analysis of JPEG Quality Factor for Image Compression Efficiency and Structural Quality

Danurkholis[1], Rafi Abhirama[1], Ginan Surya Irawan[1], Muhammad Akbar Alghifari[1], Muh Rayhan Putra Akil[1], Agustien Mulyantini[1]
[1]Department of Computer Engineering, School of Electrical Engineering, Telkom University, Indonesia

Article Info

Article history:

Received October 23, 2026
Revised November 14, 2026
Accepted November 15, 2026

Keywords:

JPEG
Quality Factor
Bits per Pixel
PSNR
SSIM
MS-SSIM
Rate–Quality Analysis

ABSTRACT

This study examined 24 Kodak Lossless True Color photographs. It investigated how compression can reduce file size while keeping the photograph clear. Using the Pillow library with 4:2:0 chroma subsampling the study tested JPEG compression at five quality factor levels: 10 30, 50, 70 and 90. Other image formats that were also examined were WebP and AVIF. To judge how good the photographs looked the study used PSNR, SSIM and MS-SSIM metrics. To evaluate how effective compression was, the study measured file size, bits per pixel and compression ratio. The findings show that as the JPEG quality factor increases, the file size increases but the photograph quality also improves. PSNR rose from 26.67 to 38.03 dB SSIM increased from 0.745 to 0.962. The average file size grew from 15.68 KiB at quality factor 10 to 112.62 KiB at quality factor 90. As the quality factor rises further the gains in photograph quality become smaller. There is a trade-off, between photograph quality and space saved at quality factor 50 and quality factor 70. Depending on whether storage or photograph quality's more important these results can help select the right JPEG compression setting. No human observers reviewed the photographs.

This is an open access article under the CC BY-SA license.

Corresponding Author:

Agustien Mulyantini
Department of Computer Engineering
School of Electrical Engineering, Telkom University
Bandung, Indonesia
Email: agustienmulyantiniam@telkomuniversity.ac.id

1. INTRODUCTION

Efficient use of storage and transmission resources is essential to wide-scale application of digital images; for example, in multimedia systems, on-line services, digital science, and archives. A large part of why so many applications use digital image compression based on the JPEG standard is because of the large volume of digital images used. The JPEG image compression standard utilizes Discrete Cosine Transform (DCT) to compress the image data. After the DCT has been applied to the image, Quantization is used to reduce the number of bits required to store the image. Finally, Entropy Coding is used to further reduce the size of the compressed file. There are several problems associated with Quantization, including but not limited to, Blocking, Blurring, and other undesirable effects [1], [2], [3].

For JPEG encoding, the quality factor (QF) defines the range of the scaling factor used in the quantization process. Low values result in smaller file sizes, but significant quality loss, while large values produce higher quality images with file size inflation [2], [4]. With QF, compression performance is not directly

controlled and image attributes, including textures and boundaries, will affect the rate of compression and the extent of degradation.

Quality assessment for images is typically done using objective metrics that require the reference image. The most common metric is the PSNR metric, which calculates the reconstruction error at the pixel level. Although easy to calculate, the PSNR metric does not always result in a metric of the quality as perceived by the viewer [5], [6]. A newer method, SSIM, calculates the image quality in the local region of the image considering luminance and the contrast [7], while the extension to multiple scales, the MS-SSIM, evaluates the image on a multi-scale level [8]. Objective metrics are computationally inexpensive but subjective evaluation is still required to capture the user's view [9].

Image compression and quality evaluation were previously researched. Santami et al. [10] utilized various objective quality metrics to evaluate JPEG vs. WebP. A two-alternative forced-choice (2AFC) observer study was performed by Afnan et al. [11] to investigate the possibility of perceptually lossless JPEG compression. Mohammadi et al. [12] performed a study using high-fidelity image compression techniques and assessed quality using JND metrics in addition to objective quality measures. Compressed-image quality assessment techniques were analyzed by Jamil [9]. Sarah et al. [13] conducted an analysis of image quality metrics by evaluating the techniques for image denoising.

More systematic reviews and analyses of the costs and benefits of increasing QF are still absent from the literature. This study aims to quantify the storage requirement of an incremental increase in QF by measuring its corresponding quality improvement. The analysis will be performed on individual images using quality measures including PSNR, SSIM, and MS-SSIM. Changes between adjacent QF values will also be examined, as well as cross-image variance.

This research investigates JPEG compression using quality factors (QF) 10, 30, 50, 70, and 90 on the Kodak Lossless True Color Image Suite. The analysis in this work is restricted to file size and objective quality metrics. As a point of reference for other compression formats, WebP and AVIF were used. Subjective observer evaluations and perceptual measures like LPIPS were excluded from the study.

2. METHOD

A quantitative, repeated-measures experimental design was used. Each of the 24 reference images was encoded at the same five QF settings. The compressed images were decoded and compared with their corresponding original RGB references. All measurements were computed per image before summary statistics and paired tests were applied.

2.1 Dataset and Uncompressed RGB Baseline

The experiments used the Kodak Lossless True Color Image Suite, consisting of 24 PNG images with dimensions of 768 × 512 or 512 × 768 pixels and 8-bit RGB representation [14]. The PNG files served only as reference pixel sources; their compressed file sizes were not used as the compression baseline. The baseline was the theoretical size of an uncompressed 24-bit RGB image:

$$F_o = \mathrm{W} \times \mathrm{H} \times 3 \tag{1}$$

where F_0 is the uncompressed RGB size in bytes, W and H are image width and height in pixels, and the factor 3 represents three 8-bit color channels. For each Kodak image, W × H = 768 × 512 = 393,216 pixels; therefore, F_o = 1,179,648 bytes, or 1,152 KiB. All 24 images were verified to have this baseline size.

2.2 JPEG Encoding Procedure

The experiments used Python 3.13 with Pillow 11.3.0 [15] along with scikit image 0.25.2 [16], NumPy 2.4.6, SciPy 1.16.3, pandas 2.2.3, and Matplotlib 3.10.0. Encoded files were decoded to RGB images for checking quality. Subsampling was looked at in the output files. There was no cropping or resizing done and dimensions stayed the same as the references.

Quantization stayed with the defaults from the quality parameter in Pillow. The QF values were chosen to show low, intermediate, and high settings over a wide range. It seems like they do not cover every option and are not equal steps in perception. Findings apply only to this Pillow setup. In total 120 JPEG files were made from 24 images across the 5 settings. File sizes include headers and are in kibibytes where one KiB is 1024 bytes. That part about the number of images times settings adds up to the total.

2.3 Evaluation Metrics

Let I_0 represent the original RGB image, I^c the decoded compressed image, and F^c the compressed file size in bytes. All full-reference quality metrics were computed by comparing I_0 with I^c.

2.3.1 Bit Rate and Size Reduction

The bit rate in bits per pixel (bpp) and size reduction relative to the uncompressed RGB baseline were calculated as:

$$bpp = \frac{(\mathrm{Fc}\ x\ 8)}{(\mathrm{W}\ \times\ \mathrm{H})} \tag{2}$$

$$\mathrm{R} = \frac{(\mathrm{F_o} - \mathrm{F_c})}{\mathrm{F_o}} \tag{3}$$

The two measures were computed for every image and QF setting. Since all images have the same pixel count, the mean per-image bpp equals the bpp calculated from the mean file size. Reduction values describe the difference from the theoretical uncompressed RGB representation, not from the PNG source files.

2.3.2 PSNR

Pixel-wise reconstruction error was measured using mean squared error over all RGB channels:

$$MSE = \frac{1}{3WH}\sum_{c=1}^{3}\sum_{x=1}^{W}\sum_{y=1}^{H}\left[I_c^{(o)}(x,y) - I_c^{(r)}(x,y)\right]^2 \tag{4}$$

$$PSNR = 10\log_{10}\left(\frac{MAX_I^2}{MSE}\right), MAX_I = 2^8 - 1 = 255 \tag{5}$$

PSNR was computed using scikit-image with data_range=255 [16]. Higher values indicate lower reconstruction error but do not necessarily reflect perceptual equivalence [5], [6].

2.3.3 SSIM

SSIM compares local luminance, contrast, and structural information [7]. It was computed on RGB images using scikit-image [16] with the following settings: win_size=None, gaussian_weights=True, use_sample_covariance=False, channel_axis=−1, and data_range=255. With gaussian_weights=True and the library defaults, the effective Gaussian window follows the 11 × 11, σ = 1.5 configuration. SSIM was evaluated over the RGB channels and averaged by metric implementation. The theoretical SSIM range is −1 to 1; all values observed in this experiment were positive, using $K_1 = 0.01$, $K_2 = 0.03$, and $L = 255$.

SSIM compares local luminance, contrast, and structural information [7]. It was computed on RGB images using scikit-image [16] with the following settings: win_size=None, gaussian_weights=True, use_sample_covariance=False, channel_axis=−1, and data_range=255. With gaussian_weights=True and the library defaults, the effective Gaussian window follows the 11 × 11, σ = 1.5 configuration. SSIM was evaluated over the RGB channels and averaged by metric implementation. The theoretical SSIM range is −1 to 1; all values observed in this experiment were positive, using $K_1 = 0.01$, $K_2 = 0.03$, and $L = 255$.

2.3.4 MS-SSIM

The MS-SSIM calculation was performed across 5 scales using the weights provided in [8], i.e., 0.0448, 0.2856, 0.3001, 0.2363, and 0.1333. A Gaussian filter with σ = 1.5 and truncation = 3.5 was utilized to estimate the local mean, variance, and covariance for each step. The contrast/structure components of MS-SSIM were evaluated for the first 4 scales, while the luminance component was evaluated at the coarsest scale. Down-sampling between scales was accomplished by performing a 2 x 2 mean pooling operation. At each scale, the average of the local value was obtained excluding a 5-pixel border. The score for each RGB channel was computed individually and then averaged.

We also compared our results with the single-scale version of SSIM available in scikit-image. On a test image, we found that the results obtained from both methods were identical up to 10-6. Additionally, we demonstrated that identical images yield a value of 1.0. However, these validations do not demonstrate that our implementation is equivalent to any other independently implemented version of MS-SSIM. One limitation of our work is that we did not compute LPIPS.it is mentioned only as a candidate perceptual metric for future work [17].

2.4 Statistical and Marginal-Efficiency Analysis

For each QF, the mean and sample standard deviation were calculated across 24 images. The Friedman test was applied separately to PSNR, SSIM, and MS-SSIM to assess differences across the five related QF conditions, with image identity treated as the block. Two-sided Wilcoxon signed-rank tests were then applied

to each consecutive QF pair (10 vs. 30, 30 vs. 50, 50 vs. 70, and 70 vs. 90). Holm adjustment was applied to the four pairwise comparisons within each metric using Friedman [18], Wilcoxon [19], and Holm [20].

Marginal efficiency was defined as the change in the mean quality metric divided by the change in mean file size between consecutive QF settings:

$$E_m = \frac{\Delta Q_m}{\Delta F} \quad \text{(quality-metric gain per additional KiB)} \tag{6}$$

where ΔQ_m is the change in mean PSNR, SSIM, or MS-SSIM and ΔF is the change in mean compressed file size in KiB. This measure is used to describe the amount of additional objective quality associated with additional storage. It is not a subjective utility function and does not establish a universal optimum.

Image-level variability was visualized using boxplots. Three visual examples were selected automatically from the QF 70 results: the image with the smallest file, the median file, and the largest file. For each selected image, a 96 × 96-pixel region of interest (ROI) was selected as the window with the highest variance of the per-pixel arithmetic mean of the three RGB channels in the original image. The same ROI coordinates were applied to the original and all five JPEG reconstructions. These visual examples are illustrative only and do not constitute a formal observer study.

2.5 Additional Codec Comparison

The additional comparison encoded the same 24 images as lossy WebP [21] and AVIF [22] using Pillow [15]; both codecs were supported in the recorded environment. Both WebP and AVIF support were detected. For these formats, the encoder was called with the nominal quality parameter set to the same five numerical settings (10, 30, 50, 70, and 90); no additional codec-specific encoding options were set. The resulting files were decoded and converted to RGB, and file size, bpp, PSNR, SSIM, and the authors' MS-SSIM implementation were calculated using the same procedures as for JPEG. Because the codecs map the nominal quality parameter to different internal encoding decisions, equal numerical quality settings are not treated as equivalent quality levels. This comparison is descriptive and limited to the tested encoders, environment, and Kodak images.

3. RESULTS AND DISCUSSION

This section presents the experimental results and discusses the effects of JPEG quality factor settings on compression efficiency and objective image quality.

3.1. JPEG Compression Results

Table 1 summarizes the mean file size, bit rate, reduction relative to the uncompressed RGB baseline, and full-reference quality metrics across the 24 Kodak image

Table 1. Mean JPEG compression results across the 24 Kodak images.

QF	Size (KiB)	BPP	Reduction (%)	PSNR (dB)	SSIM	MS-SSIM
10	15.68	0.327	98.64	26.67	0.745	0.894
30	31.67	0.660	97.25	30.49	0.861	0.963
50	43.47	0.906	96.23	32.17	0.896	0.977
70	59.46	1.239	94.84	33.92	0.923	0.985
90	112.62	2.346	90.22	38.03	0.962	0.993

The mean file size increased monotonically with QF, from 15.68 KiB at QF 10 to 112.62 KiB at QF 90. This corresponds to an increase from 0.327 to 2.346 bpp. Relative to the 1,152 KiB uncompressed RGB baseline, the mean size reduction decreased from 98.64% to 90.22%. The quality metrics also increased with QF. Mean PSNR rose from 26.67 to 38.03 dB, mean SSIM from 0.745 to 0.962, and mean MS-SSIM from 0.894 to 0.993. The metrics describe different aspects of reconstruction: PSNR reflects pixel-wise error, SSIM reflects local structural similarity, and MS-SSIM aggregates structural information across scales. Their numerical ranges and sensitivities differ, so the values should not be interpreted as interchangeable measures. Figure 1 shows file size and SSIM versus QF, and Figure 2 shows the rate–quality curves.

3.2 Rate Quality Relationship and Marginal Efficiency

The rate–quality curves show that increasing QF requires progressively more storage to obtain higher objective quality. The relevant practical question is therefore not only whether quality increases, but how much additional quality is obtained for each additional unit of storage. Table 2 reports the marginal changes between consecutive QF settings.

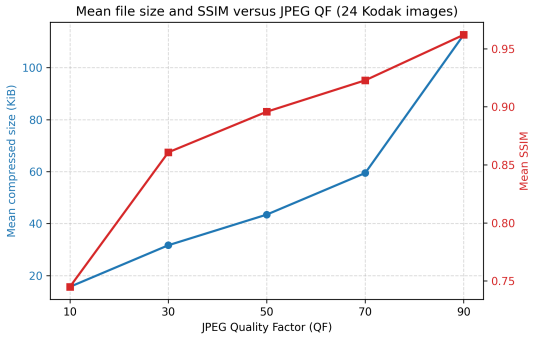

Figure 1. Mean compressed file size (left axis) and mean SSIM (right axis) versus JPEG QF across 24 Kodak images.

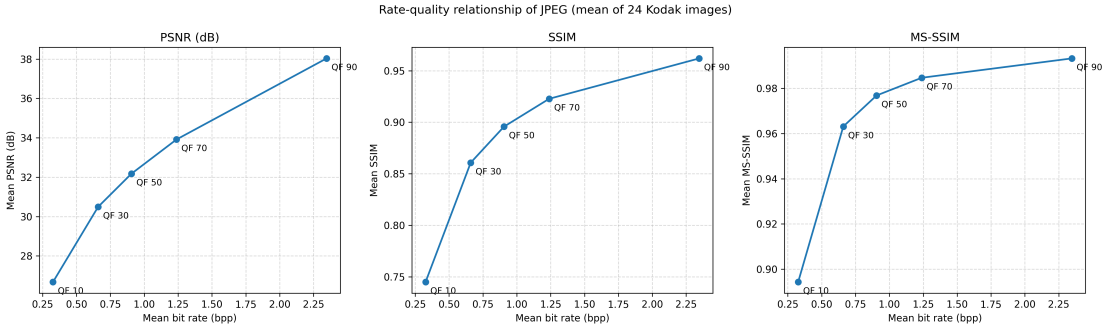

Figure 2. JPEG rate–quality curves showing mean PSNR, SSIM, and MS-SSIM against mean bit rate.

Table 2. Marginal quality gain per additional file-size unit between consecutive QF settings.

Transition	ΔSize (KiB)	ΔBPP	ΔPSNR (dB)	ΔSSIM	ΔMS-SSIM	ΔPSNR/KiB	ΔSSIM/KiB	ΔMS-SSIM/KiB
QF 10→30	15.99	0.333	3.82	0.116	0.069	0.239	0.0072	0.0043
QF 30→50	11.79	0.246	1.68	0.035	0.014	0.143	0.0030	0.0012
QF 50→70	15.99	0.333	1.74	0.027	0.008	0.109	0.0017	0.0005
QF 70→90	53.16	1.108	4.12	0.039	0.009	0.077	0.0007	0.0002

Table 2 shows that the return on additional storage declined steadily. The SSIM gain per KiB decreased from 0.0072 (QF 10→30) to 0.0030, 0.0017, and 0.0007 (QF 70→90), and the MS-SSIM gain per KiB from 0.0043 to 0.0002. The step from QF 70 to QF 90 required 53.16 KiB, about 3.3 times the 16.00 KiB needed from QF 50 to QF 70, for an SSIM increase of 0.039 versus 0.027. QF 50 (0.906 bpp, SSIM 0.896) and QF 70 (1.239 bpp, SSIM 0.923; 94.84% reduction) are therefore reported as two operating points: moving from QF 50 to QF 70 costs 0.333 bpp and yields +1.74 dB and +0.027 SSIM, whereas the next step is considerably more expensive. QF 30 is the most efficient step but reaches a lower SSIM of 0.861. These are conditional candidates, not a universal optimum.

3.3 Comparative Rate–Quality Results for JPEG, WebP, and AVIF

This research put JPEG, WebP, and AVIF head-to-head. The test bed? Twenty-four Kodak images. We ran them through five nominal quality settings each. Table 3 lays out the results: mean file size, bit rate, PSNR, SSIM, and MS-SSIM. Size reduction figures are there too. But this isn't a perfectly controlled experiment. Those identical nominal quality values don't actually map to equivalent settings across codecs. So we're looking at a descriptive comparison.

In the PSNR and SSIM panels of Figure 3, the WebP and AVIF curves lie above the JPEG curve over the overlapping bit-rate range, whereas the MS-SSIM curves converge at higher bit rates. WebP at quality 50 used 0.722 bpp versus 0.906 bpp for JPEG at QF 50 (about 20% fewer bits) with higher PSNR (33.24 vs 32.17 dB) and SSIM (0.907 vs 0.896), and AVIF at quality 50 used 0.675 bpp versus 1.239 bpp for JPEG at QF 70

(about 45% fewer bits) with comparable PSNR (34.19 vs 33.92 dB) and SSIM (0.920 vs 0.923). These comparisons are descriptive: the nominal quality values are not equivalent, default encoder settings were used, and only one dataset was tested. Figure 3 illustrates the rate–quality trajectories of the three codecs across different bit rates and objective quality levels. The curves show distinct performance patterns. However, because default settings were used and nominal quality values are not equivalent, the results represent descriptive comparisons rather than perceptually matched evaluations.

Table 3. Mean rate and objective-quality results for JPEG, WebP, and AVIF across 24 Kodak images.

Codec	Quality	Size (KiB)	BPP	Reduction (%)	PSNR (dB)	SSIM	MS-SSIM
JPEG	10	15.68	0.327	98.64	26.67	0.745	0.894
JPEG	30	31.67	0.660	97.25	30.49	0.861	0.963
JPEG	50	43.47	0.906	96.23	32.17	0.896	0.977
JPEG	70	59.46	1.239	94.84	33.92	0.923	0.985
JPEG	90	112.62	2.346	90.22	38.03	0.962	0.993
WebP	10	14.22	0.296	98.77	29.15	0.807	0.940
WebP	30	24.61	0.513	97.86	31.44	0.873	0.964
WebP	50	34.65	0.722	96.99	33.24	0.907	0.975
WebP	70	44.85	0.934	96.11	34.69	0.928	0.982
WebP	90	96.50	2.010	91.62	39.56	0.971	0.993
AVIF	10	4.38	0.091	99.62	26.32	0.704	0.882
AVIF	30	13.79	0.287	98.80	30.20	0.839	0.951
AVIF	50	32.41	0.675	97.19	34.19	0.920	0.979
AVIF	70	61.58	1.283	94.65	37.87	0.958	0.990
AVIF	90	127.46	2.655	88.94	41.77	0.983	0.996

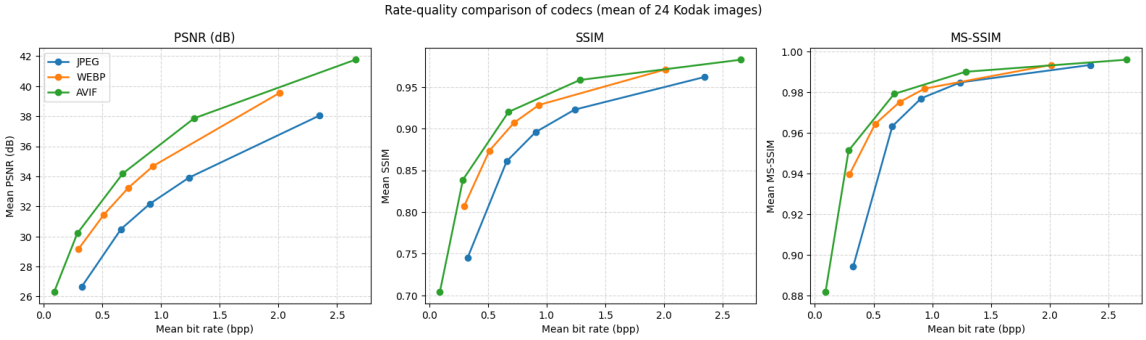

Figure 3. Rate–quality comparison of JPEG, WebP, and AVIF: mean PSNR, SSIM, and MS-SSIM versus mean bit rate (bpp). Nominal quality values are not equivalent across encoders.

3.4. Visual Examples and ROI Inspection

Figure 4 presents the original images and their JPEG reconstructions at the five QF settings for three images selected from the QF 70 file-size distribution. The selected images were kodim23, kodim19, and kodim13, representing the smallest, median, and largest QF 70 files, respectively.

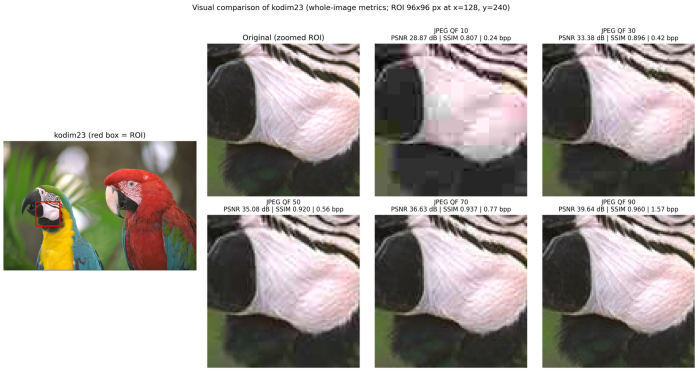

Figure 4a. Visual comparison for kodim23: original image and JPEG reconstructions at QF 10, 30, 50, 70, and 90, with the selected 96 × 96 ROI.



Figure 4b. Visual comparison for kodim19: original image and JPEG reconstructions at QF 10, 30, 50, 70, and 90, with the selected 96 × 96 ROI.

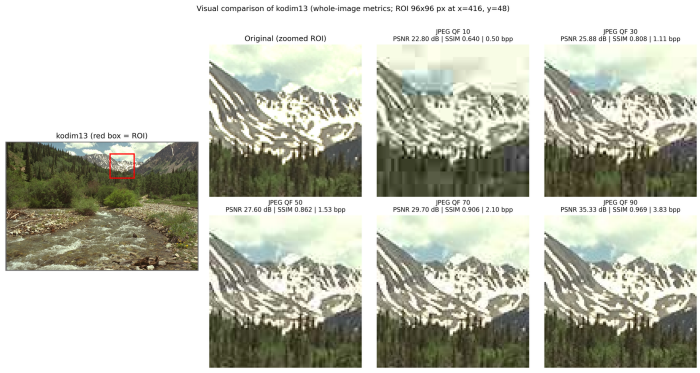

Figure 4c. Visual comparison for kodim13: original image and JPEG reconstructions at QF 10, 30, 50, 70, and 90, with the selected 96 × 96 ROI.

The enlarged regions provide a qualitative illustration of how compression artifacts vary with QF and image content. Stronger compression can produce visible block structure, smoothing of fine texture, and color degradation, while higher QF settings generally retain more local detail. The same QF does not produce identical file size or quality across images because image content affects compressibility. The ROI selection procedure is based on the spatial variance of the per-pixel arithmetic mean of the RGB channels and is intended to expose a relatively detailed region; it is not a perceptual sampling protocol. The examples therefore support visual inspection of representative outputs but do not establish a threshold of visual indistinguishability or visually lossless compression.

At QF 10, all three images show visible 8×8 block structure and loss of fine texture [3]; these artifacts are less conspicuous at QF 30 and QF 50. In kodim13, the image with the largest file at QF 70, the enlarged region still shows some smoothing of fine structure at QF 70, whereas kodim23 is closer to its original at QF 50 and QF 70. At QF 70, kodim13 reaches SSIM 0.906 and PSNR 29.70 dB at 2.10 bpp, whereas kodim23 reaches SSIM 0.937 and PSNR 36.63 dB at 0.77 bpp.

3.5 Variation Across Images and Statistical Tests

Table 4 reports the sample standard deviations across 24 images. The dispersion of file size increased from 4.28 KiB at QF 10 to 30.07 KiB at QF 90. The standard deviation of SSIM decreased from 0.049 to 0.008 as the mean SSIM approached the upper end of its range. These results indicate that storage requirements remain strongly dependent on image content, even when all images use the same encoder and QF. Image-level distributions of BPP, PSNR, SSIM, and MS-SSIM at each JPEG QF are shown in Figure 5.

Table 5 presents the Friedman and Wilcoxon signed-rank outcomes for paired image observations. The Friedman test returned $\chi^2(4) = 96.0$ with $p = 6.98 \times 10^{-20}$ across PSNR, SSIM, and MS-SSIM metrics. For each metric, every one of the 24 images showed an increase at each consecutive QF transition. The two-sided Wilcoxon signed-rank tests therefore returned the minimum attainable p-value for 24 non-zero paired differences (1.192×10^{-7}), and the Holm-adjusted value was 4.768×10^{-7} for each adjacent comparison.

The identical statistics across metrics arise because the direction of change was consistent for all images and transitions; they should not be interpreted as evidence that the magnitudes of improvement were identical.

Table 4. Sample standard deviation of compression and quality metrics across 24 images.

QF	Size SD (KiB)	BPP SD	Reduction SD (%)	PSNR SD (dB)	SSIM SD	MS-SSIM SD
10	4.28	0.089	0.37	1.79	0.049	0.022
30	9.82	0.205	0.85	2.05	0.026	0.010
50	13.45	0.280	1.17	2.03	0.019	0.006
70	17.98	0.375	1.56	1.92	0.013	0.004
90	30.07	0.626	2.61	1.44	0.008	0.002

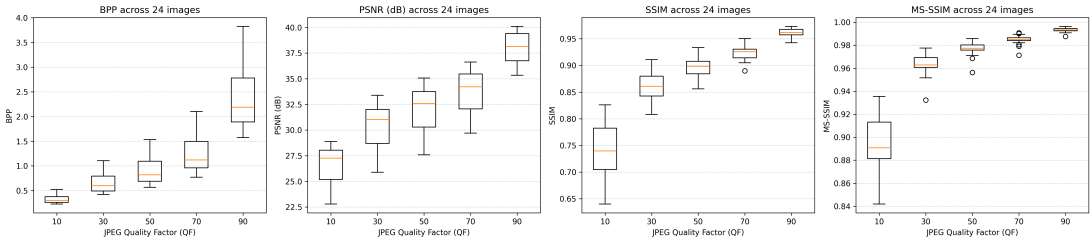

Figure 5. Image-level distributions of BPP, PSNR, SSIM, and MS-SSIM at each JPEG QF.

Table 5. Friedman and Wilcoxon signed-rank results for the paired image observations (N = 24).

Metric	Friedman χ^2 (4)	Friedman p	Wilcoxon raw p (each adjacent pair)	Holm adjusted p	Images increasing
PSNR	96.00	6.98×10^{-20}	1.192×10^{-7}	4.768×10^{-7}	24/24 at each transition
SSIM	96.00	6.98×10^{-20}	1.192×10^{-7}	4.768×10^{-7}	24/24 at each transition
MS-SSIM	96.00	6.98×10^{-20}	1.192×10^{-7}	4.768×10^{-7}	24/24 at each transition

The paired tests establish that the direction of the quality trend is consistent across the tested images. They do not quantify the practical value of each increment; that question is addressed by the marginal-efficiency analysis. The results also do not imply that all images have similar storage costs or that a particular QF is optimal for every image.

3.6 Scope of the Empirical Contribution

The results should be seen as an empirical look at how much storage is needed and how good the image quality is when using one specific JPEG setup.It's not meant to be a new way of compressing images, a new way to measure quality, or a general comparison of all image formats. The contribution is in putting together an uncompressed RGB base, bit-rate reports, image-based multi-metric tests, efficiency analysis, and a detailed comparison of different codecs, all on the same group of 24 images. The operating points and code paths seen are unique to this dataset and the encoder settings that were tested

4. CONCLUSION

This study evaluated how the JPEG quality factor (QF) relates to storage cost and objective reconstruction quality using 24 images from the Kodak Lossless True Color Image Suite. JPEG was encoded at five QF settings with Pillow (4:2:0 chroma subsampling, optimize=False). Relative to the uncompressed RGB baseline of 1,152 KiB per image, mean file size increased from 15.68 KiB (0.327 bpp; 98.64% reduction) at QF 10 to 112.62 KiB (2.346 bpp; 90.22% reduction) at QF 90, while mean PSNR increased from 26.67 to 38.03 dB, SSIM from 0.745 to 0.962, and MS-SSIM from 0.894 to 0.993. All three metrics increased at every consecutive QF transition for all 24 images (Holm-adjusted $p < 0.001$).

Marginal analysis showed diminishing returns: the SSIM gain per KiB fell from 0.0072 (QF 10–30) to 0.0007 (QF 70–90). QF 50 (0.906 bpp, SSIM 0.896) and QF 70 (1.239 bpp, SSIM 0.923) are therefore

conditional operating points: QF 50 offers lower storage use, whereas QF 70 provides higher measured fidelity for 0.333 bpp more. Neither is a universal optimum, because file size and quality vary across images at the same QF and the appropriate choice depends on the application's storage and quality requirements. In the additional comparison, WebP produced smaller mean files than JPEG at all five nominal settings, whereas

AVIF reached the highest objective scores at nominal quality 90 at the largest mean file size (127.46 KiB). These comparisons are descriptive; nominal quality settings are not equivalent across codecs.

The study is limited to one 24-image dataset, five nominal quality settings, and default Pillow settings for WebP and AVIF. The custom MS-SSIM implementation was not cross-validated against an independent canonical implementation. No subjective assessment or LPIPS evaluation [17] was performed, so the results do not establish visually lossless compression. Future work should test additional datasets and encoder implementations, use a denser quality grid, validate the MS-SSIM implementation independently, and conduct controlled observer studies and comparisons at matched bit rates or matched subjective quality.

5. DATA AND CODE AVAILABILITY

The source code, analysis notebook, processed measurements, and experimental results supporting this study are publicly available at https://github.com/Dannn-116/jpeg-qf-tradeoff-kodak (archived at https://doi.org/10.5281/zenodo.24191856).

REFERENCES

[1] G. K. Wallace, "The JPEG still picture compression standard," IEEE Trans. Consumer Electron., vol. 38, no. 1, pp. xviii–xxxiv, Feb. 1992, doi: 10.1109/30.125072.
[2] ITU-T, "Information technology—Digital compression and coding of continuous-tone still images—Requirements and guidelines," Recommendation T.81, Sep. 1992.
[3] S. A. Golestaneh and D. M. Chandler, "Algorithm for JPEG artifact reduction via local edge regeneration," J. Electron. Imaging, vol. 23, no. 1, Art. no. 013018, Feb. 2014, doi: 10.1117/1.JEI.23.1.013018.
[4] Independent JPEG Group, "libjpeg." [Online]. Available: https://www.ijg.org/.
[5] Z. Wang and A. C. Bovik, "Mean squared error: Love it or leave it? A new look at signal fidelity measures," IEEE Signal Process. Mag., vol. 26, no. 1, pp. 98–117, Jan. 2009, doi: 10.1109/MSP.2008.930649.
[6] Q. Huynh-Thu and M. Ghanbari, "Scope of validity of PSNR in image/video quality assessment," Electron. Lett., vol. 44, no. 13, pp. 800–801, Jun. 2008, doi: 10.1049/el:20080522.
[7] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," IEEE Trans. Image Process., vol. 13, no. 4, pp. 600–612, Apr. 2004, doi: 10.1109/TIP.2003.819861.
[8] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in Conf. Rec. 37th Asilomar Conf. Signals, Syst. Comput., vol. 2, 2003, pp. 1398–1402, doi: 10.1109/ACSSC.2003.1292216.
[9] S. Jamil, "Review of Image Quality Assessment Methods for Compressed Images," J. Imaging, vol. 10, no. 5, Art. no. 113, May 2024, doi: 10.3390/jimaging10050113.
[10] L. D. Santami, S. R. Alamsyah, E. Irwansyah, and R. A. Charisma, "Objective Study of Image Quality Compressed with JPEG vs WebP," in Proc. 8th Int. Conf. Comput. Informatics Eng. (IC2IE), 2025, doi: 10.1109/IC2IE67206.2025.11283342.
[11] A. Afnan, F. Ullah, Y. Yaseen, J. Lee, S. Jamil, and O.-J. Kwon, "Subjective Assessment of Objective Image Quality Metrics Range Guaranteeing Visually Lossless Compression," Sensors, vol. 23, no. 3, Art. no. 1297, 2023, doi: 10.3390/s23031297.
[12] S. Mohammadi, M. Jenadeleh, J. Sneyers, D. Saupe, and J. Ascenso, "Evaluation of Objective Image Quality Metrics for High-Fidelity Image Compression," IEEE Access, vol. 14, pp. 35651–35668, 2026, doi: 10.1109/ACCESS.2026.3669417.
[13] U. Sara, M. Akter, and M. S. Uddin, "Image Quality Assessment through FSIM, SSIM, MSE and PSNR—A Comparative Study," J. Comput. Commun., vol. 7, no. 3, pp. 8–18, Mar. 2019, doi: 10.4236/jcc.2019.73002.
[14] Eastman Kodak Company, "Kodak Lossless True Color Image Suite." [Online]. Available: https://r0k.us/graphics/kodak/. Accessed: Nov. 13, 2026.
[15] Pillow Contributors, "Pillow documentation." [Online]. Available: https://pillow.readthedocs.io/.
[16] S. van der Walt et al., "scikit-image: image processing in Python," PeerJ, vol. 2, Art. no. e453, Jun. 2014, doi: 10.7717/peerj.453.
[17] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2018, pp. 586–595, doi: 10.1109/CVPR.2018.00068.
[18] M. Friedman, "The use of ranks to avoid the assumption of normality implicit in the analysis of variance," J. Amer. Statist. Assoc., vol. 32, no. 200, pp. 675–701, Dec. 1937, doi:

10.1080/01621459.1937.10503522.
[19] F. Wilcoxon, "Individual comparisons by ranking methods," Biometrics Bull., vol. 1, no. 6, pp. 80–83, Dec. 1945, doi: 10.2307/3001968.
[20] S. Holm, "A simple sequentially rejective multiple test procedure," Scand. J. Statist., vol. 6, no. 2, pp. 65–70, 1979.
[21] Google, "WebP: An image format for the Web." [Online]. Available: https://developers.google.com/speed/webp.
[22] Alliance for Open Media, "AV1 Image File Format (AVIF)," AVIF Specification v1.0.0. [Online]. Available: https://aomediacodec.github.io/av1-avif/.