Optimal JPEG Quantization Table Fusion by Optimizing Texture Mosaic Images and Predicting Textures

Chen-Hsiu Huang (✉ chenhsiu48@cmlab.csie.ntu.edu.tw)

National Taiwan University <https://orcid.org/0000-0003-4352-0375>

Ja-Ling Wu

National Taiwan University

Research

Keywords: JPEG Quantization Table, Simulated Annealing, Texture Prediction

Posted Date: June 17th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-604349/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

RESEARCH

Optimal JPEG Quantization Table Fusion by Optimizing Texture Mosaic Images and Predicting Textures

Chen-Hsiu Huang* and Ja-Ling Wu

*Correspondence:
chenhsu48@cmlab.csie.ntu.edu.tw
Computer Science and Information
Engineering, National Taiwan
University, 106 Taipei, Taiwan
Full list of author information is
available at the end of the article

Abstract

The JPEG standard allows the use of a customized quantization table; however, it is still challenging to find an optimal quantization table timely. This work aims to solve the dilemma of balancing computational cost and image-specific optimality by introducing a new concept of texture mosaic images. Instead of optimizing a single image or a collection of representative images, the conventional JPEG optimization techniques can be applied to the texture mosaic image to obtain an optimal quantization table for each texture category. We use the simulated annealing technique as an example to validate our framework. To effectively learn the visual features of textures, we use the ImageNet pre-trained MobileNetV2 model to train and predict the new image's texture distribution, then fuse optimal texture tables to come out with an image-specific optimal quantization table. Our experiment demonstrates around 30% size reduction with a slight decrease of FSIM quality but visually indistinguishable on the evaluation datasets. Moreover, our rate-distortion curve shows superior and competitive performance against other prior works under a high-quality setting. The proposed method, denoted as JQF, achieves per image optimality for JPEG encoding with less than one second additional timing cost.

Keywords: JPEG Quantization Table; Simulated Annealing; Texture Prediction

Introduction

As of today, JPEG [3] remains the most used lossy image compression standard for digital images in content sharing and various image capture devices [1] [2]. Since JPEG was developed in 1992, new image coding standards such as JPEG 2000 [4], WebP [5], and HEIF [6] have been proposed to respond to the continual expansion of multimedia applications. The JPEG 2000 adopts the Discrete Wavelet Transform (DWT) and arithmetic coding with a layered file format that offers flexibility. Compared to JPEG's fixed-size transform, the DWT is in principle open-ended for image size and compresses better than JPEG. The WebP and HEIF leverage the intra-frame coding technologies from video codecs. The WebP is based on the VP8 coding standard, uses intra-prediction, in-loop filtering, and block adaptive quantization technologies to win over JPEG. The HEIF uses the still image coding of the state-of-the-art High Efficiency Video Coding (HEVC) standard [7], which is roughly double the video compression efficiency compared to the Advanced Video Coding (AVC) standard [8].

Although the new formats are proven to be superior to JPEG subjectively and objectively [9] [10], JPEG remains the most used image format in the past three

decades. The JPEG 2000 is broadly used in medical images [11] and remote sensing images in Geographic Information Systems (GIS) [12]. The WebP image format is advocated by Google and is well supported in popular web browsers. The WebP has become the second popular Internet image format next to the JPEG [1], but merely a 1.2% market share. Since 2017, Apple switches to use HEIF as the default photo capturing format in iPhone [13]. We expect more mobile phone manufacturers to follow, and the digital imaging world is one step closer to replace JPEG with the state-of-the-art HEVC [6] coded image format. However, we expect JPEG to stay for at least a few more decades [2] because JPEG still satisfies average user demands. As JPEG takes the top-1 Internet image traffic of 72.9%, many kinds of research on how to reduce the JPEG bandwidth have been proposed by tech giants like Google [14] [15] [16] and Facebook [17].

The success of JPEG comes from its low computational complexity and coding effectiveness considering the human visual system (HVS). JPEG typically achieves 10:1 compression gain without noticeable quality loss and is very cheap to implement. JPEG divides the image into 8×8 blocks, using Discrete Cosine Transform (DCT) to shift the pixels from the spatial domain to the frequency domain. The transformed DCT coefficients are rearranged in zigzag order. A quantization table is used to quantize DCT coefficients, resulting in reduced coefficients and sparse DCT block for run-length encoding (RLE). Figure 1 (a) shows the default luminance quantization table provided in the JPEG standard.

Since JPEG is lossy, the compression rate can be adjusted, allowing a selectable tradeoff between storage size and image quality. Although a smaller quality metric Q achieves better image size reduction, it may introduce visual artifacts such as blocking and ringing effects. Therefore, imaging applications usually set the default JPEG quality at 75 or above. The JPEG allows the use of customized tables, but how to optimize the quantization table remains challenging due to the numerous solution space and lack of reliable HVS quality measurements. As a result, modern image applications and digital camera processors tend to compress JPEG images with minimal quantization values to preserve quality.

Related Works

Rate-distortion Optimization

The JPEG standard table is obtained from a series of psychovisual experiments, which determines the visibility thresholds for the DCT basis functions at pre-defined

viewing distance [2]. Since early 90's, there are many previous works to explore the potential of designing better quantization tables [19] [20] [21] [22] [23] [24] [25] [26] [27] [28] [29]. Some of the works directly focus on improve rate-distortion performance and differ mainly in the optimization strategy, such as coordinate-descent enumeration [21] [26] [30], dynamic programming [22], and genetic algorithms [23].

Early works such as Ratnakar and Livny [22] uses Mean Squared Error (MSE) as the distortion measurement during optimization, which has been proven poorly correlated to HVS. Ernawan and Nugraini [25] proposed a psychovisual threshold through quantitative experiments based on the contribution of DCT coefficients on each frequency order to the reconstruction error. Their psychovisual threshold assigns an optimal balance between quality and compression ratio and derives a new quantization table using 40 real images. Significant performance improvement is reported, but it seems not generally applicable to all images.

Hu et al. [27] indicate that human perceived distortion is discretely characterized by some jumps and could be modeled with a stair quality function (SQF). The SQF utilizes the notion of just noticeable difference (JND), which refers to the visibility threshold below any change that can not be detected by the HVS [31]. With the JND concept, Zhang et al. [28] redefine the distortion metric by soft-thresholding MSE values according to the JND as a threshold for each coefficient and propose a JND-based perceptual optimization method. Hu's method operates as image-specific optimization on the testing image with an exhaustive search algorithm. Their work delivers significant bitrate saving compared to JPEG standard and shows superior size reduction vs. quality tradeoff than Ernawan's PSY table. However, Hu et al. do not report the time cost for the image-specific optimization.

In recent years, many researchers proposed deep learning-based methods to end-to-end optimize the JPEG quantization table for image compression [32], artifacts correction [17], and specific computer vision tasks [33] [16]. Two essential issues, non-differentiable JPEG quantization and entropy coding estimation, have been appropriately handled. The end-to-end learning for the quantization table with neural network backpropagation is achieved.

Simulated Annealing

The stochastic optimization process known as simulated annealing [34] has successfully been applied to find vector quantization parameters. Monro and Sherlock [19] [20] attempted to use simulated annealing to determine quantization tables for DCT coding. Monro et al. anneal all the 64 quantization values with the cost function composed of RMSE error and a selected target compression ratio. Their optimization process searches optimal tables on selected images with minimal RMSE error while keeping the chosen target's compression ratio. Around a single-digit percentage improvement of the RMSE error is reported comparing to the standard JPEG table.

Until the early 2000s, new objective FR-IQA methods like SSIM [35] and FSIM [36] were proposed and shown to be statistically closer to HVS. Jiang et al. [24] utilize SSIM as the quality metric to evaluate image distortion during the simulated annealing. In their work, a multi-objective optimization equation is proposed to minimize bitrate while maximizing SSIM. To solve the equation, they estimate

the Pareto optimal point for finding an optimal quantization table so that no other feasible points have a lower bitrate and higher SSIM index. However, since Pareto optimal point differs in every image, the multi-objective optimization framework only proves helpful on a per-image basis. Hopkins et al. [29] adopt FSIM as the quality metric and revise the annealing process to focus on compression maximization with a temperature function that rewards lower error. They use the standard JPEG table as the initial table and randomly perturb ten quantization table values at each step. A set of 4,000 images was selected from the RAISE [37] dataset as a training set to run four groups of annealing processes at quality metrics 35, 50, 75, and 95. Four global optimized quantization tables are reported to reduce the compressed size by around 20% over the JPEG standard table and claimed to improve FSIM error by 10%. However, the corpus of 4,000 training images is merely a proxy to the universal pictures but not custom-tailored per image.

Image-Specific Optimization

Another work from Google's JPEG encoder Guetzli, [14] aims to produce visually indistinguishable images at a reduced bitstream size. Using a close-loop optimizer, Guetzli optimizes global quantization tables and selectively zeros out specific DCT coefficients in each block. However, the most size reduction of Guetzli comes from identifying DCT coefficients to zero out without significantly decrease the quality score. Guetzli's approach is too aggressive to zero out DCT coefficients and causes noticeable artifacts at a low bitrate setting. Compared to Hopkins' global optimized quantization table, Guetzli's per image optimization strategy achieves a 29-45% data size reduction at the cost of computational complexity and memory consumption (up to 30 minutes on a high-resolution image), which is considered impractical in real-world applications.

The SJPEG encoder [39] uses a simple adaptive quantization method by collecting and analyzing the DCT coefficients' histogram in each sub-band. Starts with the scaled standard table, the SJPEG uses a least-square fitted slope to decide an overall λ value for the rate-distortion cost function. With the obtained λ , SJPEG searches for the best quantizer with the minimal cost within each sub-band's delta range. The rate and distortion are approximated with the logarithm of quantized coefficient and square of quantization error. The adaptive quantization method is fast because it simplifies the optimization problem and uses quantization error as the quality measurement.

Proposed Method

Figure 2 shows the block diagram of the proposed JPEG Quantization Table Fusion (JQF) framework. The training workflow is a series of offline procedures to cluster texture patches and optimize texture quantization tables. The prediction workflow splits the input image into patches of the same size as training, predicted by the deep CNN model to obtain a texture distribution, which structurally describes what kinds of texture categories compose the input image. Then we can further aggregate those optimal tables to form a custom-tailored quantization table. We use the simulated annealing technique to validate our framework. Other conventional rate-distortion based [21] [26] [30] [22] [23], perceptual based [27] [28], or deep learning-based [17] [33] [16] JPEG optimization methods can be used as well.

Figure 2 The workflows of the proposed JQF method. (a) The texture training and mosaic image annealing flow. (b) The texture prediction and quantization table fusion flow.

Texture Patches Clustering

Since the JPEG uses 8×8 image blocks for DCT transform, the texture patch size B must be a multiplier of 8. We choose block size $B = 64$ to crop the training images into 64×64 texture patches in a non-overlapping manner. Because we do not know how many types of textures in our RAISE training set, the unsupervised K-Means clustering algorithm is used to cluster textures into $K = 100$ categories.

Typically, the image clustering problem can be considered two separated steps: 1) design the appropriate visual features to extract and 2) train a logistic regression classifier that minimizes the class assignments [40]. In recent years, the success of deep Convolutional Neural Networks (CNNs) in various computer vision tasks [41] [42] makes CNN a preferred choice for feature extraction. The ImageNet [41] pre-trained CNN models such as VGG-16 [42] and MobileNetV2 [43] have become the building blocks in many computer vision applications. In this context, we use the last activation maps after layers of ConvNets, called bottleneck features, to extract texture features. In our experiences, both the pre-trained VGG-16 and the MobileNetV2 model could extract representative visual features for clustering. We choose the MobileNetV2 network for its compact model size and prediction efficiency on mobile devices. To speed up the K-Means clustering algorithm, we apply the principal component analysis (PCA) on the extracted 1280-dimensional feature vectors to further reduce the dimension to 500, covering 82% of the variance. Figure 3 (a) and 3 (b) are two examples of texture mosaic images from the RAISE training set.

Annealing on Texture Mosaic Images

The quantization table is an 8×8 matrix of 8-bit unsigned values. It is unrealistic to enumerate the whole space for an optimal solution. Therefore, we anneal the JPEG's

default luminance quantization table in Figure 1 (a) of the texture mosaic images with quality metric $Q = 95$ to validate our method. Assuming that luminance affects visual quality the most, we use the fixed standard chrominance table during the annealing process. In each step, we randomly choose some table indices based on the magnitude of the current value. We weigh smaller values more, assuming that smaller value has higher visual importance than others. Then we update the table values with ± 1 in probability to obtain a candidate solution. To comply with the JPEG reference implementations, we scale the candidate solution to target quality metrics and check if it is updated in the scaled table or randomly perturb values again. We compress a new JPEG file using the candidate table as the luminance table and standard chrominance table. If the new JPEG file has size reduction and the quality degradation falls within tolerance $\gamma = 0.005$, we accept the candidate solution and complete the current iteration. We evaluate the FSIM quality reduction by,

$$\text{FSIM}(I_r, I_c) \geq \text{FSIM}(I_r, I_s) \times (1 - \gamma), \quad (1)$$

where I_r is the raw image, I_c and I_s are the JPEG images compressed with the candidate table and standard table. The physical meaning of equation (1) is that we will only accept a solution with FSIM degradation of less than 0.5%. To not be trapped in a local minimum, there is a probability $P(i)$ to accept a worse solution, affected by the temperature function $T(i)$ and the energy delta ΔE from the current solution score improvement ratio. The temperature function used to affect the probability is given by,

$$T(i) = \frac{M}{M + i \times \rho} \quad \text{for } i = 1 \dots M \quad (2)$$

where i is the iteration index, M is the maximum iterations to anneal, ρ is a constant to control the temperature function so that the probability approaches $\frac{1}{\rho+1}$ at the end of the annealing process. The probability $P(i)$ to receive an answer

is calculated by,

$$P(i) = \Delta E \times T(i) \quad (3)$$

$$\Delta E = \frac{S_i}{S_{i-1}} \quad (4)$$

$$S_i = C_i \times (1 - D_i) \quad (5)$$

where S_i is the current solution's score, which is composed of C_i as the current compressed JPEG file size, and $D_i = \text{FSIM}(I_r, I_c)$ as the quality distortion.

With the design of probability $P(i)$, we have a higher possibility to accept a worse solution at the early stage, which prevents us from being trapped in a local minimum. Furthermore, the probability decreases gradually with the number of iterations we run; then, the annealing becomes a hill-climbing process. The current iteration could be finished by accepting a worse solution or randomly updating again to get the next candidate. In this work, we choose $M = 3000$ and $\rho = 10$ to anneal each texture mosaic image t and obtain corresponding optimized table $\mathcal{T}_t$.

We show the final annealed quantization tables $\mathcal{T}_1$ and $\mathcal{T}_{19}$ of categories id 1 and 19 in Figure 3. Table $\mathcal{T}_1$ and $\mathcal{T}_{19}$ obtain a size reduction of 37.9% and 25.42% with the FSIM quality degradation of 0.48% and 0.4%, but they are visually indistinguishable. We conclude two findings by observing the annealed results: 1) a smoothed texture like category 1 could achieve more bitrate saving than a complex texture with high-frequency signals under the given quality distortion tolerance. 2) the annealing process tends to raise the quantizer magnitude in low-frequency bands and reduce quantizers in high-frequency bands.

Texture Training and Prediction

We use the pre-trained MobileNetV2 network to extract texture features for clustering; we fine-tune the ConvNets parameters and train the fully connected layers for K texture categories. The texture patches from the RAISE dataset are split into 80%-20% for training and testing. We employ the Adam optimizer with default settings in PyTorch with a learning rate of 0.0001 and batch size 2048. We train the texture CNN model for 30 epochs, resulting in top-3 testing accuracy reaches 95.08%. We crop the input image into same-sized patches for prediction and then execute the forward propagation pass to obtain each patch's texture category. Figure 4 shows the texture distribution of the lighthouse (kodim19) image from the Kodak dataset; textures are displayed using their texture mosaic images.

Quantization Table Fusion

We consider two strategies to aggregate texture optimized quantization tables, voting by the majority and weighted average. The voting strategy will pick the texture category's quantization table with the most counts in the histogram as the final quantization table. The weighted average strategy weights the annealed quantizers by each texture category's percentage. We select the weighted average policy for its overall better performance. The fused optimal quantization table $\mathcal{T}_O$ is calculated by,

$$\mathcal{T}_O(i) = \sum_t (\mathcal{T}_t(i) \times W_t) \quad \text{for } i = 0 \dots 63 \quad (6)$$

where $\mathcal{T}_t$ denotes the optimal table of the texture t , and W_t is the weights of the corresponding textures. We provide the final fused quantization table at $Q = 95$ of the lighthouse image in Figure 1 (b). The fused table has the same tendency as we observed when annealing texture tables, i.e., larger quantizers in low-frequency bands and smaller quantizers in high-frequency. Our observation echoes the conclusions obtained from Monro and Sherlock [19, 20] that the JPEG standard table improperly over-estimates the low-frequency parts and under-estimates the high frequencies.

Experimental Results

Dataset and Implementation Details

We use the RAISE dataset [37] as our training database and cross-validate it with the Kodak dataset [44]. The RAISE dataset is a real-world image dataset primarily designed for the evaluation of digital forgery detection algorithms. It consists of 8,156 high-resolution raw images in resolution around 4288×2848 , uncompressed, and guaranteed to be camera-native. The Kodak dataset [44] has 24 lossless images, commonly used for evaluating image compression. Each image is about 768×512 in resolution.

We randomly select 50 images from the RAISE-1k subset as the testing set. The rest of the images are used as the training set. We crop the 1,487 training images into $B \times B$ patches with stride 256, generating a total of 261,712 texture patches. Then we cluster patches to K textures and perform the simulated annealing process on the stitched mosaic images. At most 225 patches are randomly selected from each texture category to limit the required annealing time. We compare our JQF method's performance with the JPEG standard table (denotes as STD), Hopkins' proposed global table [29] (denotes as SAT), and the SJPEG [39]. Since both JQF and SAT optimize the luminance table while SJPEG's optimization includes the chrominance table, we encode all the raw images in grayscale JPEG as a fair comparison. The compression is done with the command line program `cjpeg` from `libjpeg` [18] and the `sjpeg` command.

Baseline Performance

Table 1 shows the baseline performance of block size $B = 64$, $K = 100$ texture categories, $M = 3000$ annealing iterations, and $Q = 95$ targeting JPEG quality.

Table 1 The baseline performance with $B=64$, $K=100$, $M=3000$, $Q=95$.

Dataset		RAISE	Kodak
JQF v.s. SAT	Size	-15.58%	-13.48%
	PSNR	0.64%	1.11%
	SSIM	-0.02%	0.00%
	FSIM	-0.07%	-0.03%
JQF v.s. Standard	Size	-32.50%	-28.34%
	PSNR	-4.39%	-4.22%
	SSIM	-0.76%	-0.68%
	FSIM	-0.40%	-0.47%
JQF v.s. SJPEG	Size	2.22%	1.06%
	PSNR	0.97%	0.39%
	SSIM	0.40%	0.22%
	FSIM	0.13%	0.12%
SJPEG v.s. SAT	Size	-17.28%	-14.34%
	PSNR	-0.32%	0.72%
	SSIM	-0.43%	-0.22%
	FSIM	-0.20%	-0.15%

The proposed annealing and fusion framework for the two benchmark datasets delivers fairly close FSIM quality to the SAT, decreasing by 0.07% and 0.03%, but it further reduces the JPEG compressed file size by 15.58% and 13.48%. In other words, compared with the JPEG standard, our method achieves a significant size reduction (near 30%) with a slight (0.4%) FSIM quality decrease. The FSIM quality degradation falls within our annealing criterion $\gamma = 0.5\%$ and is visually indistinguishable, as shown in Figure 5. Interestingly, our method improves the PSNR by 0.64% and 1.11% against SAT. Although it is well known that PSNR does not align with the HVS, it reflects the signal fidelity. We do not mind having a higher PSNR if we can reduce the compressed size. Compared to SJPEG, our proposed method is less aggressive for the same testing datasets with 2.22% and 1.06% size increases but with overall higher qualities. Both SJPEG and JQF have improved PSNR against SAT on the Kodak dataset; we think it indicates that the image-specific optimization better adapts to the individual image than a global one.

Ablation Study

Annealing Iterations

The simulated annealing algorithm terminates with an optimal global solution with probability one as long as the iteration number is large [45]. However, finding suitable iterations to ensure a significant probability of success is time-consuming. Instead, we choose to run annealing iterations that yield similar FSIM to that of Hopkins' work as a fair control parameter. The Table 2 shows the result. More iterations to anneal could reduce the compressed file size while keeping the average FSIM quality degradation within the $\gamma = 0.5\%$ tolerance. Since the -0.4% FSIM reduction is close to our preset tolerance, we choose $M=3000$ as our default annealing steps.

Quantization Table Fusion

We compare the optimization performance of the $K=100$ texture mosaic images with the 50 RAISE testing images in Table 3. The 3000-iteration annealing process

delivers a 26.26% size reduction compared to the standard table on those 100 texture images. However, the weighted average of those optimal texture tables could further boost the size reduction to 29.97% on real-world testing images. The result shows that the fused quantization table does better adapt to the image content.

Texture Block Size

Table 4 shows the size reduction comparisons on block sizes 32, 48, 64, and 96, respectively, using 100 texture clusters. The outcome aligns with our expectation that smaller-sized image blocks generally adapt to image content better, so the overall size reduction tendency is apparent. However, the improvement is not significant, except for the more considerable performance drop with block size 96. It is intuitive because a large image block could cover specific objects and cannot represent certain textures, or the texture prediction accuracy drops.

Number of Texture Categories

The optimization performances of category numbers $K=25$, 50, 100, and 150 are shown in Table 5. Not to our surprise that the more texture categories we split, the

Table 2 The performance comparison of annealing iterations on RAISE dataset, with B=64, K=100, Q=95.

Iterations	JQF v.s. Standard			
	Size	PSNR	SSIM	FSIM
M=2000	-25.14%	-3.10%	-0.52%	-0.28%
M=3000	-29.97%	-4.46%	-0.79%	-0.40%

Table 3 The performance comparison of RAISE texture mosaic images and RAISE testing images, with B=64, K=100, M=3000, Q=95.

RAISE Images		100 Textures	50 Testings
JQF v.s. Standard	Size	-26.26%	-29.97%
	PSNR	-4.95%	-4.46%
	SSIM	-0.52%	-0.79%
	FSIM	-0.46%	-0.40%

better resize reduction we can achieve. Still, the improvement tendency is noticeable but not significant when we split into more categories. Theoretically, we can cluster image patches into a massive set of textures to describe the image content. Nevertheless, since our annealing process runs on those texture mosaic images, annealing on an enormous collection of images may become impractical.

Rate-Distortion Curve Comparison

We compare our JQF method with JPEG standard (STD), Hopkins' simulated annealing table [45] (SAT), SJPEG [39], and Ernawan's psychovisual table [25] (PSY), and Luo's deep learning derived tables [16] (DLT). We use SJPEG to encode images from the testing set with the quality setting $Q=35, 75, 50$, and 95 , then obtain the corresponding averaging bitrate to generate four rate-distortion data points as the benchmark. We adjust the JPEG quality target Q for each competing method to generate the closest bitrate and record its quality metrics. The rate-distortion curve comparison of dataset Kodak and RAISE is shown in Figure 6.

Firstly, we observed the PSY table has its SSIM and FSIM performance very close to the JPEG standard but slightly higher PSNR. That makes sense because PSY derived the new quantization table according to a primitive threshold from psychovisual experiments, which helps maintain the HVS quality metrics.

Secondly, the SAT table does not come out right size vs. quality tradeoff, even compared to the JPEG standard. Both SAT and our method run simulated annealing techniques with FSIM error and compressed size as the acceptance criteria, so the two approaches will generate inferior FSIM rate-distortion performance. However, the SAT table seems unable to gain more signal fidelity in PSNR, especially in the RAISE dataset.

The third point, our annealing process uses the FSIM as an error budget to exchange for better size reduction. The stochastic process explores the solution space to search for a more optimized table. Furthermore, the texture-based quantization fusion approach further reduced the distortion and enhanced the PSNR RD-curve performance. Our method and SJPEG deliver significant PSNR gain at JPEG's high-quality settings as an image-specific optimization approach. The proposed JQF method performs less optimally than JPEG at low-quality settings because we use the same annealing iterations $M=3000$ for all targeting quality. If we anneal fewer iterations or relax the tolerance γ in equation (1), we will obtain optimal texture tables with better quality but worse compression ratio.

Table 4 The performance comparison of different block size on RAISE dataset, with K=100, M=3000, Q=95.

Block Size		B=32	B=48	B=64	B=96
JQF v.s. Standard	Size	-30.62%	-30.58%	-29.97%	-28.14%
	PSNR	-4.64%	-4.63%	-4.46%	-3.96%
	SSIM	-0.84%	-0.82%	-0.79%	-0.67%
	FSIM	-0.41%	-0.42%	-0.40%	-0.36%

Table 5 The performance comparison of texture categories on RAISE dataset, with B=64, M=3000, Q=95.

Texture Categories		K=25	K=50	K=100	K=150
JQF v.s. Standard	Size	-29.71%	-29.87%	-29.97%	-30.16%
	PSNR	-4.46%	-4.52%	-4.46%	-4.47%
	SSIM	-0.78%	-0.80%	-0.79%	-0.79%
	FSIM	-0.42%	-0.40%	-0.40%	-0.41%

Figure 6 The rate-distortion curve comparison conducted on the Kodak dataset and RAISE testing set. The proposed JQF method, SJPEG, and DLT are in the winning group in terms of PSNR coding efficiency but cannot maintain the SSIM and FSIM quality level. Especially DLT, the deep learning method has the worst SSIM and FSIM score under low bitrate.

At last, the JQF, SJPEG, and DLT are in the winning group and performed relatively close to each other. The quality of our proposed JQF is slightly ahead of that of the SJPEG. Luo et al. demonstrate the advantage of end-to-end learning to optimize a global optimized table that beats image-specific optimization approaches like JQF and SJPEG. The lead is not significant, but it is surprising. Their deep learning method can obtain another 0.6dB PSNR gain if applied to specific images, but the computational analysis is not included in their paper [16]. On the SSIM and FSIM HVS-oriented metrics, JQF and SAT provide lower performance curves because simulated annealing methods trade the quality loss for the bitrate reduction. SJPEG approximates its visual quality using DCT coefficient rounding error to trade the perceptual quality for speed; therefore, it also provides lower SSIM and FSIM curves. The situation is the same in DLT, which only formulates pixel reconstruction error in the cost function, producing the worst HVS metric in quality.

Prediction Computational Cost

We compare the time cost of the proposed JQF method and SJPEG adaptive quantization in Table 6. The MobileNetV2-based prediction model is light-weight in size (only 10.5MB) and efficient to compute. Furthermore, for high-resolution images, we can predict the texture distribution on a down-sampled version. We used a workstation with Intel Core i7-9700K CPU and Nvidia GeForce RTX 2080 Ti GPU for the benchmark, taking about 0.11 seconds with GPU and 1.55 seconds on pure CPU computation per RAISE testing image. The computing cost of JQF is competitive to SJPEG with GPU and acceptable with CPU. Given that the deep learning inference chips may become more generally available in the future, predicting texture distribution for JPEG optimization remains feasible.

Discussion

Figure 7 demonstrates the visual artifacts introduced by all competing methods. The artifacts are positively correlated with the bitrate savings at low bitrate ($Q=35$).

Table 6 The time cost comparison of quantization table optimization methods, in seconds.

Method	Dataset	GPU Time	CPU Time
JQF	RAISE	0.1115	1.5545
	Kodak	0.0060	0.1842
SJPEG	RAISE	0.2950	0.2950
	Kodak	0.0100	0.0100

As suggested in [29] [14], we think the JPEG optimization problem finds its value at a high quality setting, where there is more room for optimization to trade human unperceivable quality for bitrate saving. As a 30-years-old codec, JPEG merely applies predictive coding on DC coefficients after DCT transform, and the coding gain mainly comes from consecutive quantized zero coefficients for RLE. There is no intra-prediction coding tool like HEVC [6] in JPEG, so the optimization at low bitrate becomes a zero-sum game of size or quality. For low bitrate JPEG optimization, there are alternative research works being proposed to either pre-editing [15] or post-processing with artifacts correction [17] to maintain good visual quality at reduced bitrate.

Conclusion

We propose a novel JPEG quantization table fusion (JQF) framework to solve the dilemma of balancing between computational cost and image-specific optimality by introducing a new concept of texture mosaic images. We use the simulated annealing technique as a proof-of-concept to validate our method. We use the K-Means clustering algorithm to cluster texture categories and the MobileNetV2 pre-trained CNN model to learn and predict textures. Then we fuse an optimal texture table from texture-optimized tables to come out with an image-specific optimal quantization table. Our experiments demonstrate a significant size reduction but visually indistinguishable in perceptual quality compared with that of the JPEG standard table on the Kodak dataset and a real-world consumer photo dataset. Our method's rate-distortion curve shows superior and competitive performance against other prior works under JPEG high bitrate setting. With the proposed JQF framework, the per-image optimality for JPEG encoding is achieved with affordable additional computational cost.

Declarations

Acknowledgements

The authors would like to thank the anonymous reviewers for their helpful comments.

Funding

This research is supported by the MOST of Taiwan and CITI SINICA, under the grant numbers MOST 108-2218-e-002-055, MOST 108-2221-E-002-103-my3, and Sinica 3012-C3447.

Availability of data and materials

The code for the quantization table fusion is available at <https://github.com/chenhsiu48/JQFv2>.

Competing interests

The authors declare that they have no competing interests.

Abbreviations

HEIF: High Efficiency Image File; SSIM: Structural Similarity Index Measure; FSIM: Feature Similarity Index Measure; HVS: Human Visual System; FR-IQA: Full-Reference Image Quality Assessment.

Authors' contributions

Chen-Hsiu Huang contributed to the design and implementation of the research, to the analysis of the results and to the writing of the manuscript. Ja-Ling Wu supervised the project.

Authors' information

Chen-Hsiu Huang received the B.S. and M.S. degrees in computer science from the Department of Computer Science and Information Engineering (CSIE), National Taiwan University (NTU), in 2002 and 2004. He is now pursuing the Ph.D. degree with the Department of CSIE, NTU.

Ja-Ling Wu has been a professor with the Department of Computer Science and Information Engineering, National Taiwan University (NTU), since 1996. From 2004 to 2007, he was the Head of the Graduate Institute of Networking and Multimedia, NTU. He was selected to be one of the lifetime Distinguished Professors of NTU, in 2006.

Author details

Computer Science and Information Engineering, National Taiwan University, 106 Taipei, Taiwan.

References

- Usage statistics of image file formats for websites. https://w3techs.com/technologies/overview/image_format
- Hudson, G., Léger, A., Niss, B., Sebestyén, I., Vaaben, J.: Jpeg-1 standard 25 years: past, present, and future reasons for a success. *Journal of Electronic Imaging* **27**(4), 040901 (2018)
- Wallace, G.K.: The jpeg still picture compression standard. *IEEE transactions on consumer electronics* **38**(1), (1992)
- Skodras, A., Christopoulos, C., Ebrahimi, T.: The jpeg 2000 still image compression standard. *IEEE Signal processing magazine* **18**(5), 36–58 (2001)
- WebP: A new image format for the Web. <https://developers.google.com/speed/webp>
- Lainema, J., Hannuksela, M.M., Vadakital, V.K.M., Aksu, E.B.: Hvc still image coding and high efficiency image file format. In: 2016 IEEE International Conference on Image Processing (ICIP), pp. 71–75 (2016). IEEE
- Sector, I.T.S., Ito, O.: Itu-t recommendation h. 265: high efficiency video coding. Telecommunication Standardization Sector, Geneva, Switzerland (2013)
- Ohm, J.-R., Sullivan, G.J., Schwarz, H., Tan, T.K., Wiegand, T.: Comparison of the coding efficiency of video coding standards—including high efficiency video coding (hevc). *IEEE Transactions on circuits and systems for video technology* **22**(12), 1669–1684 (2012)
- Nguyen, T., Marpe, D.: Objective performance evaluation of the hevc main still picture profile. *IEEE Transactions on circuits and systems for video technology* **25**(5), 790–797 (2014)
- Hanhart, P., Rerabek, M., Korshunov, P., Ebrahimi, T.: Subjective evaluation of hevc intra coding for still image compression. In: Seventh International Workshop on Video Processing and Quality Metrics for Consumer Electronics-VPQM 2013 (2013)
- Przelaskowski, A.: The jpeg2000 standard for medical image applications. *Task Quarterly* **8**(2), 147–158 (2004)
- Brower, B., Fiete, R., Shuler, R.: Jpeg 2000 application in gis and remote sensing. In: The JPEG 2000 Suite, pp. 305–339. John Wiley & Sons, Ltd Chichester, UK, ??? (2009)
- Snell, J.: iOS 11: HEVC, HEIF, and What You Need to Know About These New Video and Photo Formats. <https://www.macworld.com/article/3226490/> Accessed 2021-02-08
- Alakuijala, J., Obryk, R., Stoliarchuk, O., Szabadka, Z., Vandevenne, L., Wassenberg, J.: Guetzli: Perceptually guided jpeg encoder. *arXiv preprint arXiv:1703.04421* (2017)
- Talebi, H., Kelly, D., Luo, X., Dorado, I.G., Yang, F., Milanfar, P., Elad, M.: Better compression with deep pre-editing. *arXiv preprint arXiv:2002.00113* (2020)
- Luo, X., Talebi, H., Yang, F., Elad, M., Milanfar, P.: The rate-distortion-accuracy tradeoff: Jpeg case study. *arXiv preprint arXiv:2008.00605* (2020)
- Ehrlich, M., Davis, L., Lim, S.-N., Shrivastava, A.: Quantization guided jpeg artifact correction. *Proceedings of the European Conference on Computer Vision* (2020)
- Group, I.J., et al.: Libjpeg. <https://www.ijg.org/>
- Monro, D.M., Sherlock, B.G.: Optimum dct quantization. In: [Proceedings] DCC93: Data Compression Conference, pp. 188–194 (1993). IEEE
- Sherlock, B., Nagpal, A., Monro, D.: A model for jpeg quantization. In: Proceedings of ICSIPNN'94. International Conference on Speech, Image Processing and Neural Networks, pp. 176–179 (1994). IEEE
- Wu, S.-W., Gersho, A.: Rate-constrained picture-adaptive quantization for jpeg baseline coders. In: 1993 IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 5, pp. 389–392 (1993). IEEE
- Ratnakar, V., Livny, M.: Rd-opt: An efficient algorithm for optimizing dct quantization tables. In: Proceedings DCC'95 Data Compression Conference, pp. 332–341 (1995). IEEE
- Costa, L.F., Veiga, A.C.P.: Identification of the best quantization table using genetic algorithms. In: PACRIM. 2005 IEEE Pacific Rim Conference on Communications, Computers and Signal Processing, 2005., pp. 570–573 (2005). IEEE
- Jiang, Y., Pattichis, M.S.: Jpeg image compression using quantization table optimization based on perceptual image quality assessment. In: 2011 Conference Record of the Forty Fifth Asilomar Conference on Signals, Systems and Computers (ASILOMAR), pp. 225–229 (2011). IEEE
- Ernawan, F., Nugraini, S.H.: The optimal quantization matrices for jpeg image compression from psychovisual threshold. *Journal of Theoretical and Applied Information Technology* **70**(3), 566–572 (2014)
- Fung, H.T., Parker, K.J.: Design of image-adaptive quantization tables for jpeg. *Journal of Electronic Imaging* **4**(2), 144–150 (1995)
- Hu, S., Wang, H., Kuo, C.-C.J.: A gmm-based stair quality model for human perceived jpeg images. In: 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 1070–1074 (2016). IEEE

28. Zhang, X., Wang, S., Gu, K., Lin, W., Ma, S., Gao, W.: Just-noticeable difference-based perceptual optimization for jpeg compression. *IEEE Signal Processing Letters* **24**(1), 96–100 (2016)
29. Hopkins, M., Mitzenmacher, M., Wagner-Carena, S.: Simulated annealing for jpeg quantization. *arXiv preprint arXiv:1709.00649* (2017)
30. Yang, E.-h., Wang, L.: Joint optimization of run-length coding, huffman coding, and quantization table with complete baseline jpeg decoder compatibility. *IEEE Transactions on Image Processing* **18**(1), 63–74 (2008)
31. Jayant, N., Johnston, J., Safranek, R.: Signal compression based on models of human perception. *Proceedings of the IEEE* **81**(10), 1385–1422 (1993)
32. Strümpler, Y., Yang, R., Timofte, R.: Learning to improve image compression without changing the standard decoder. *arXiv preprint arXiv:2009.12927* (2020)
33. Choi, J., Han, B.: Task-aware quantization network for jpeg image compression. In: *European Conference on Computer Vision*, pp. 309–324 (2020). Springer
34. Kirkpatrick, S., Gelatt, C.D., Vecchi, M.P.: Optimization by simulated annealing. *science* **220**(4598), 671–680 (1983)
35. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing* **13**(4), 600–612 (2004)
36. Zhang, L., Zhang, L., Mou, X., Zhang, D.: Fsim: A feature similarity index for image quality assessment. *IEEE transactions on Image Processing* **20**(8), 2378–2386 (2011)
37. Dang-Nguyen, D.-T., Pasquini, C., Conotter, V., Boato, G.: Raise: a raw images dataset for digital image forensics. In: *Proceedings of the 6th ACM Multimedia Systems Conference*, pp. 219–224 (2015)
38. Alakuijala, J., et al.: Butteraugli. <https://github.com/google/butteraugli> (2016)
39. The sjpeg library. <https://github.com/webmproject/sjpeg>
40. Caron, M., Bojanowski, P., Joulin, A., Douze, M.: Deep clustering for unsupervised learning of visual features. In: *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 132–149 (2018)
41. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: *Advances in Neural Information Processing Systems*, pp. 1097–1105 (2012)
42. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014)
43. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L.-C.: Mobilenetv2: Inverted residuals and linear bottlenecks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4510–4520 (2018)
44. Kodak PhotoCD dataset. <http://r0k.us/graphics/kodak/>
45. Granville, V., Krivánek, M., Rasson, J.-P.: Simulated annealing: A proof of convergence. *IEEE transactions on pattern analysis and machine intelligence* **16**(6), 652–656 (1994)