DIFFERENTIABLE JPEG-GUIDED PRE-PROCESSING FOR PERCEPTUAL ENHANCEMENT AT EXTREMELY LOW BITRATES

Wei-Jhih Huang, Yi-Chen Yen**

Department of Communication Engineering, National Central University, Taiwan

ABSTRACT

JPEG remains a dominant image compression standard but introduces blocking and banding artifacts at low bitrates. Decoder-side deep-learning enhancement improves quality but complicates deployment and compatibility. We instead propose a differentiable JPEG-guided pre-processing framework that learns to produce compression-resilient images without modifying the decoder. By combining a robust differentiable JPEG simulator with a color-reduction network, our method enables stable end-to-end training and effectively compensates quantization artifacts. Experiments show consistently improved perceptual quality over standard JPEG and prior pre-processing approaches.

Index Terms— image compression, differentiable JPEG, pre-processing, artifact reduction.

1. INTRODUCTION

The ubiquitous JPEG standard exhibits severe blocking and banding artifacts at low quality factors [1]. While deep learning post-processing can mitigate these artifacts, it typically requires computationally expensive decoder-side networks and hardware changes. Consequently, pre-processing approaches have gained traction for their backward compatibility and ease of deployment. Shoda et al. [2] proposed a color reduction method to suppress noise, but end-to-end optimization is hindered by the non-differentiable nature of JPEG quantization. To enable gradient-based learning, Shin et al. [3] introduced an initial differentiable approximation of the rounding function, though its accuracy and training stability remain limited.

In this work, we propose a pre-processing framework that enhances low-bitrate JPEG images by driving a color reduction network [2] with the precise differentiable JPEG simulator of Reich et al. [4]. By replacing unstable approximations [3] with a model that rigorously accounts for quantization scaling and clipping, we enable stable end-to-end optimization. Our contributions are: (1) the integration of a robust differentiable simulator [4] to resolve training instability inherent in prior methods; (2) an optimized perceptual loss configuration that effectively suppresses blocking artifacts while preventing gradient vanishing; and (3) superior visual quality at low bitrates compared to standard baselines, achieved without modifying the decoder.

2. THE METHOD

We propose an end-to-end framework that integrates a color reduction network with a robust differentiable JPEG simulator. We adopt the backbone utilizing a Gaussian Mixture Model for palette extraction and a U-Net for residual estimation. A critical deviation in our approach is the exclusion of the heuristic pseudo-contour suppression module used in prior studies. We observed that such signal-processing steps often introduce structural degradation; instead, we rely entirely on the neural network to learn artifact-suppressing features in a data-driven manner.

To enable stable end-to-end optimization, we replace unstable approximations with a robust differentiable JPEG simulator. Previous implementations often neglected precise input value ranges and quantization table scaling, leading to gradient vanishing. Our simulator strictly models the discrete operations of the standard codec, including differentiable clipping to enforce the $[0, 255]$ pixel range. This ensures the input to the quantization step matches the dynamic range of standard JPEG, resolving convergence instability and allowing the network to effectively learn from compression artifacts.

To suppress color banding and blocking artifacts typical of low-bitrate compression ($Q = 10$), we employ a composite loss function emphasizing perceptual quality. We assign a dominant weight to the perceptual component, as it forces the network to generate residuals that specifically counteract visual discontinuities rather than merely minimizing pixel-wise errors. The total objective is defined as:

$$\mathcal{L} = \lambda_1 \mathcal{L}_1 + \lambda_{vgg} \mathcal{L}_{VGG} + \lambda_{lpips} \mathcal{L}_{LPIPS} \tag{1}$$

3. EXPERIMENTAL RESULTS

We trained the model on DIV2K for 50 epochs with a learning rate of 10^{-4} and a training quality factor of $Q = 10$, using a composite perceptual loss with weights $\lambda_1 = 0.8$, $\lambda_{vgg} = 0.01$, and $\lambda_{lpips} = 0.4$. Evaluation is performed

* Shared first authorship and contributed equally to this work.

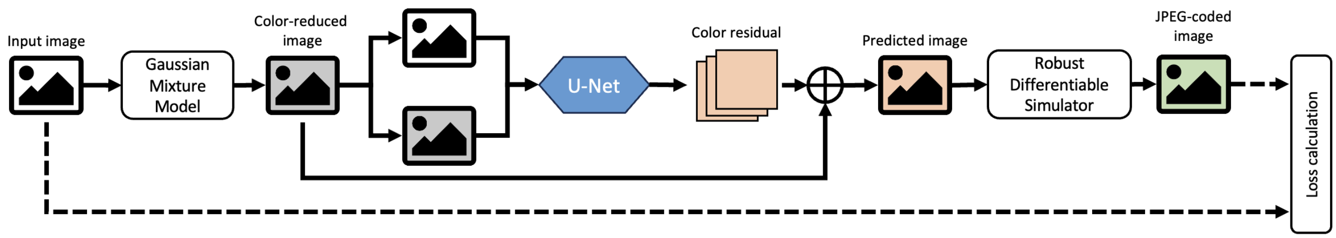

Fig. 1. Our Proposed Architecture

Table 1. Quantitative comparison at standard compression level ($Q = 10$). Higher PSNR/SSIM and lower LPIPS indicate better quality.

Method	PSNR ↑	SSIM ↑	LPIPS ↓
Standard JPEG	**26.3808**	**0.7699**	0.4602
Shoda et al. [2]	23.3616	0.7051	0.5219
Shin et al. [3]	23.2460	0.7266	0.4663
Proposed Method (Ours)	25.2663	0.7593	**0.4566**

on 16 test images at $Q = 10$ and $Q = 1$. We compare our method with Standard JPEG, the preprocessing approach of Shoda et al. [2] (built on the differentiable approximation of Shin et al. [3]), and a direct implementation of Shin et al. Unlike these approximation-based baselines, our method employs the state-of-the-art differentiable JPEG simulator of Reich et al. [4] for stable end-to-end optimization. Quantitative results are shown in Table 1 and Table 2.

As expected, Standard JPEG attains the highest PSNR (26.38 dB), since any pre-processing inevitably perturbs pixel values and is penalized by MSE-based metrics. Our objective, however, is perceptual quality rather than exact pixel fidelity. As reported in Table 1, the proposed method achieves the lowest LPIPS score of *0.4566*, outperforming both Standard JPEG (0.4602) and prior learning-based approaches. This indicates that our framework more effectively suppresses blocking artifacts and low-frequency color banding, which are visually disturbing despite yielding similar PSNR. Moreover, our method consistently improves SSIM over other pre-processing baselines, highlighting the stability and effectiveness of the integrated differentiable simulator.

To further assess robustness, we evaluate the model under extreme compression ($Q = 1$). As shown in Table 2, our method surpasses Standard JPEG in both structural similarity (SSIM) and perceptual distance (LPIPS). Although PSNR is slightly lower, the preserved structure and reduced visual distortion suggest that the learned pre-compensation internalizes compression-aware priors, whereas Standard JPEG suffers from severe information loss and structural collapse at such low bitrates.

Table 2. Performance under extreme compression ($Q = 1$). Although PSNR decreases slightly, our method achieves stronger structural (SSIM) and perceptual (LPIPS) quality.

Method	PSNR ↑	SSIM ↑	LPIPS ↓
Standard JPEG	**21.3099**	0.5970	0.6273
Proposed Method (Ours)	21.0559	**0.6022**	**0.6155**

4. CONCLUSIONS

This paper integrates a color-reduction network with a differentiable JPEG simulator, replacing heuristic signal processing with a compression model to improve perceptual quality at low bitrates. Future directions include adaptive loss weighting and extension to other codecs.

5. ACKNOWLEDGMENT

The authors thank Gemini and ChatGPT for improving the grammar and word choice of their written English.

6. REFERENCES

[1] X. Fu, X. Wang, A. Liu, J. Han, and Z. Zha, "Learning dual priors for JPEG compression artifacts removal," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021, pp. 6820–6829.

[2] A. Shoda, T. Miyazaki, and S. Omachi, "JPEG Image Enhancement with Pre-Processing of Color Reduction and Smoothing," in *Proceedings of Sensors*, vol. 23, no. 21, 2023, p. 8861.

[3] R. Shin and D. Song, "JPEG-Resistant Adversarial Images," in *Proceedings of the NIPS Workshop on Machine Learning and Computer Security*, vol. 1, 2017.

[4] C. Reich, B. Debnath, D. Patel, and S. Chakradhar, "Differentiable JPEG: The Devil is in the Details," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2024, pp. 4126–4135.