This CVPR paper is the Open Access version, provided by the Computer Vision Foundation.
Except for this watermark, it is identical to the accepted version;
the final published version of the proceedings is available on IEEE Xplore.

Frequency-Biased Synergistic Design for Image Compression and Compensation

Jiaming Liu[1,†], Qi Zheng[1,†], Zihao Liu[2], Yilian Zhong[1], Peiye Liu[2], Tao Liu[3], Shusong Xu[2], Yanheng Lu[2], Sicheng Li[2], Dimin Niu[2], and Yibo Fan[1,*]

[1] State Key Laboratory of Integrated Chips and Systems, Fudan University,
[2] Damo Academy & Hupan Lab, Alibaba Group, [3] Lawrence Technological University

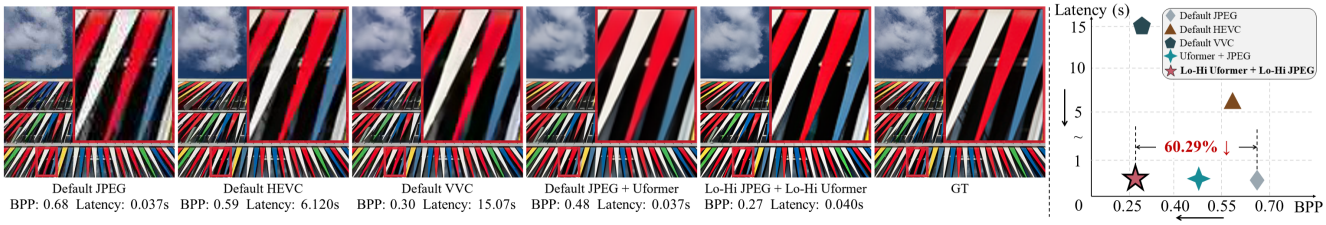


Figure 1. Comparison of five compression and/or compensation frameworks at the same LPIPS 0.164. The proposed Lo-Hi framework achieves up to 60.29% and 43.75% BPP savings compared to JPEG and SOTA NN-based CAR method, respectively. Moreover, the proposed framework achieves 376.75× and 153× latency reduction compared with VVC and HEVC, respectively.

Abstract

Compression artifacts removal (CAR), an effective post-processing method to reduce compression distortion in edge-side codecs, demonstrates remarkable results by utilizing convolutional neural networks (CNNs) on high computational power cloud side. Traditional image compression reduces redundancy in the frequency domain, and we observed that CNNs also exhibit a bias in frequency domain when handling compression distortions. However, no prior research leverages this frequency bias to design compression methods tailored to CAR CNNs, or vice versa. In this paper, we present a synergistic design that bridges the gap between image compression and learnable compensation for CAR. Our investigation reveals that different compensation networks have varying effects on low and high-frequencies. Building upon these insights, we propose a pioneering redesign of the quantization process, a fundamental component in lossy image compression, to more effectively compress low-frequency information. Additionally, we devise a novel compensation framework that applies different neural networks for reconstructing different frequencies, incorporating a basis attention block to prioritize intentionally dropped low-frequency information, thereby enhancing the overall compensation. We instantiate two compensation networks based on this synergistic design and conduct extensive experiments on three image compression standards, demonstrating that our approach significantly reduces bitrate consumption while delivering high perceptual quality.

This work was supported in part by the National Key R&D Program of China (2023YFB4502802), in part by the National Natural Science Foundation of China (62031009), in part by Fudan-ZTE joint lab, in part by Alibaba Innovative Research (AIR) Program, in part by Alibaba Research Fellow (ARF) Program. †contribute equally. * Yibo Fan is the corresponding author (fanyibo@fudan.edu.cn).

1. Introduction

The surge in mobile data has increased the need to transmit vast volumes of internet traffic from edge devices to the cloud [3]. Addressing this challenge typically involves two solutions: (1) high-quality compression at low-bitrate on the edge (2) low-quality post-compensation in the cloud [21].

At the compression stage, advanced codecs such as H.265/HEVC [28] and H.266/VVC [6], achieve lower bitrates at high perceptual quality, yet at the cost of increased computational complexity. In contrast, earlier codecs like JPEG, while computationally simpler, consume more bitrates to maintain perceptual quality, as illustrated in Fig. 1. The resource-constrained edge devices make the computationally heavy approach impractical. Thus, achieving high perceptual quality at low bitrates on resource-constrained edge devices remains challenging.

At the compensation stage, the cloud's extensive computational resources enable the use of complex post-processing methods, i.e., neural network-based CAR models [15, 17, 32], which correct distortions introduced by edge compression. However, the compensation benefits are gradually diminished by increasing computational complexity or enlarging the model size. One significant reason is that the design of compression and compensation are separate, failing to exploit the inherent connection between compression and compensation in information processing. Therefore, we aim to develop a synergistic design to fully explore potential

of the compression and compensation system.

Recent work [38] improves compression algorithms by considering the compensation network, training a compression parameter prediction network to provide block-wise parameters based on a specific restoration network. However, it does not achieve a synergistic optimization, since it remains strongly coupled to the compensation network, thus limiting the adaptability of the edge-side compressor, and the prediction network adds computational overhead, posing challenges for edge deployment. These limitations motivate us to develop a synergistic and edge-friendly design of compression and compensation stages.

We adopt this co-design approach, but in a completely distinct manner. A crucial insight enabling this design is our observation of specific frequency characteristics present in both stages—namely, frequency bias in the compensation stage and frequency selectability in the compression stage. **Frequency bias:** Recent works [16, 24] study the frequency bias of CNNs towards image restoration. We further explore such bias through a well-controlled experiment with results aligning with works [16, 24]. As shown in Fig. 2, various types of CNNs exhibit distinct effects across frequency bands, with a tendency toward better recovery of low-frequency component (FC) information than high FC. **Frequency selectability:** Image compression generally consists of three main steps—transformation, quantization, and entropy coding. The primary information loss occurs during quantization, where traditional codecs, designed with human visual system (HVS) alignment in mind, often prioritize preserving low-frequency components. Increasing the step size reduces the retained information and vice versa, enabling selective preservation of specific frequencies as required.

In summary, CNNs are more effective at compensating low-FCs, while traditional codecs are inherently optimized to preserve these low frequencies. In other words, **there remains low-frequency redundancies across existing compression-compensation frameworks.** Therefore, we aim to release the redundancies by embracing the intrinsic characteristics of compression and compensation stages in a synergistic design.

Building on the analysis, we propose a synergistic framework that combines standard image codecs with compensation neural networks. Specifically, we redesign the quantization process in standard codecs by increasing the quantization step for selected low-frequency components through a statistical-based quantization method, as these components are more easily recovered by CNNs. For the compensation network, we adopt a two-part design: a heavy-weight regression network dedicated to low-frequency compensation and a light-weight generative network focused on high-frequency restoration. And we design the base attention block (BAB) tailored to restore the low-frequency components intentionally reduced during compression. Our contributions can be summarized as follows:

1. We analyze, for the first time, the frequency bias in CNN compensation, guiding the design of the synergistic framework for compression and compensation.
2. We propose a statistical quantization, called Lo-Hi sensitive quantization, to selectively compress more low-frequency components.
3. We develop a two-step Lo-Hi compensation framework and a dedicated frequency restoration block, the base attention block (BAB), to better recover low-frequency components.
4. Comprehensive experiments demonstrate the effectiveness and efficiency of our framework, achieving the average BD-rate reduction of 45.74% and 48.09% for human vision and machine vision, respectively.

As shown in Fig. 1, compared with standard image codecs and previous compensation models, our synergistic framework achieves remarkable bitrate savings and minimal latency when reaching the given level of perceptual quality.

2. Related Works

2.1. Lossy Image Compression

Deep learning-based image compression algorithms [1, 2, 10, 29, 46] have yet to see widespread adoption on resource-constrained edge devices due to their substantial computational overhead. Instead, traditional lossy image compression standards have evolved over the past three decades to facilitate the efficient transmission and storage of edge-side image data. The inaugural lossy compression standard, proposed in 1992, is JPEG [30]. Subsequent standards include JPEG2000 [27], WebP [26], BPG (HEVC intra) [28], and VVC intra [6]. In lossy compression algorithms, the major steps causing a loss of information and leading to compression artifacts are quantization and rounding. Compression artifacts manifest as visual distortions in the reconstructed image, representing a trade-off between achieving a higher compression ratio and maintaining image quality.

2.2. Deep Learning-based CAR

To mitigate the compression artifacts, there has been a notable surge in deep learning-based CAR models. FBCNN [15] handles different quality factors for JPEG artifacts removal by embedding the quality factor derived from a decoupler module into the reconstruction module. NAFNet [9] simplifies nonlinear activation functions, saving computational costs while exhibiting competitive performance on image deblurring. Uformer [33] introduces a hierarchical encoder-decoder network consisting of Transformer blocks, incorporating local and global dependencies. CRESNet [11] leverages a multi-stage progressive network tailored for blind JPEG artifacts removal. MARN [14] incorporates a multi-scale feature extraction block to learn

salient image features, where two branches accounting for coarse and fine aspects are fused to remove JPEG image artifacts. Li et al. [17] propose a new JPEG image localization in which wide-receptive-field attention is designed to learn JPEG compression artifacts. However, few work considers compression when designing CNN model.

2.3. Perceptual Quality Metrics

PSNR measures average pixel errors yet correlating poorly with HVS [4]. NIQE [20] demonstrates high consistency with visual perception, which is also assumed to be high-frequency-aware by considering the covariance of Multivariate Gaussian distribution fit to natural scene statistics features. The naturalness and reality of authentic images can be well measured by NIQE [20]. Moreover, LPIPS [42] calculates differences in CNN activations to align with human perception, while FID [13] assesses the diversity and realism of generated content. Hence, we utilize reliable quality metrics to comprehensively evaluate lossy compressed images, including LPIPS, NIQE, FID, and PSNR.

3. Frequency Bias at Image Compensation

Numerous studies have explored the optimization learning biases of two distinct types of deep neural networks. The first type is regression models; the authors from work [22] examine the artifacts produced by regression models and delve into the spectral characteristics of CNNs throughout their training phase. The second type belongs to generative models; the authors of works [7, 24] center around the capabilities of generative models in capturing high-frequency details. Drawing inspiration from these studies, we embarked on a series of experiments to explore how frequency bias affects performance in image compensation tasks.

In our study, we initially delve into the issue of frequency bias in the task of compressed image compensation. To address this, we isolate the compensation process, framing it as a reduction problem for different Discrete Cosine Transform (DCT) FCs [12]. Assume $X(i,j)$ is an $n \times n$ image, which can be represented as the accumulation of $n \times n$ FCs of DCT:

$$X(i,j) = \sum_{u=0}^{n-1} \sum_{v=0}^{n-1} c_{(u,v)} \cdot b_{(i,j,u,v)}, \tag{1}$$

where $c_{(u,v)} \in \mathbb{R}^1$ and $b_{(i,j,u,v)} \in \mathbb{R}^{n \times n}$ represent the DCT coefficient and the corresponding basis function at different frequencies defined by (u, v), respectively.

We employ RRDBNet as the generator and UNet as the discriminator for the generative network [31]. We also use RRDBNet for the regression network. The loss functions for the generative and regression networks are $0.0001 \times L_1 + L_{GAN}$ and L_1, respectively. We randomly select 1500 training images and 200 validation images from the ImageNet [23]. We use the 8×8 DCT, dividing the

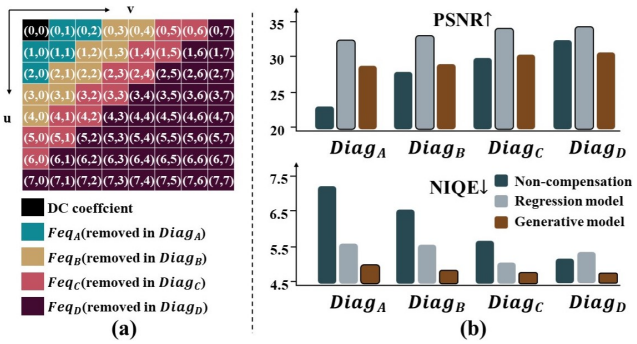


Figure 2. Datasets and results in our exploration. (a) FCs removal; (b) The compensation of regression/generative models on different FCs in PSNR and NIQE.

images into 64 FCs. As shown in Fig. 2 (a), we generate four datasets by intentionally removing FCs with large quantization steps from low (i.e., $Diag_A$) to high (i.e., $Diag_D$) in ascending order.

As shown at the top of Fig. 2 (b), for pixel-fidelity measured by PSNR, both the generative and regression models are more effective at restoring lower-frequency information than higher frequencies, and the regression model outperforms the generative model in restoring pixel errors. Beyond pixel fidelity, we also use NIQE to assess realism (refer to Sec. 2.3), as shown at the bottom of Fig. 2 (b). The NIQE scores indicate that the generative model consistently outperforms the regression model across four cases in producing realistic images, particularly in $Diag_D$, where high FCs are most reduced. However, the regression model performs even worse than the uncompensated result in $Diag_D$.

We can draw two observations: **(1) Frequency compensation bias.** Both models consistently yield superior compensation outcomes on both quality metrics in restoring low-frequency information; **(2) Model type bias.** The regression model is more adept for fidelity compensation, while the generative model excels at realism enhancement.

The *frequency compensation bias* suggests that low-frequency information can be more heavily compressed, as it is readily compensated by CNNs—contrasting with the HVS-based quantization approach in conventional codecs, which emphasizes preserving low frequencies. Therefore,

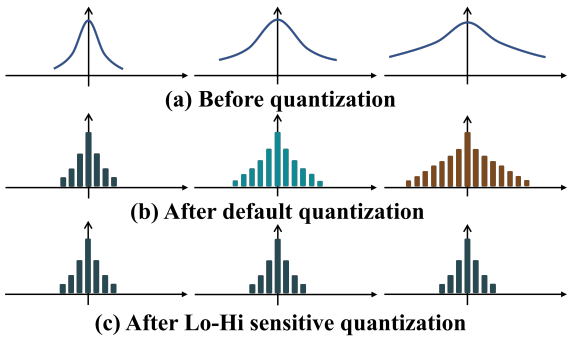


Figure 3. FC Coefficient distributions of different quantization.

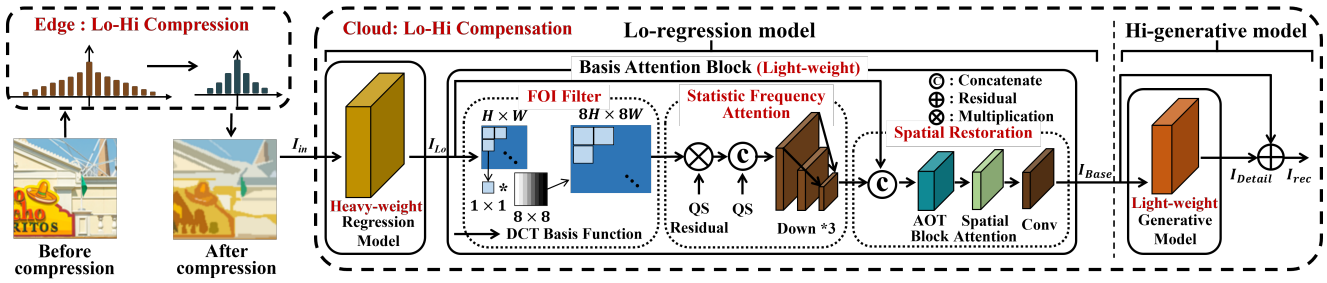


Figure 4. Proposed synergistic design and two-step Lo-Hi compensation CNN framework, including Lo-regression model and Hi-generative model. Basis attention block is designed to compensate for frequency of interest.

there remain low-frequency redundancies in previous codec and CAR frameworks. And existing CAR approaches have not addressed the *model type bias*, which involves using different CNN model types to repair distinct frequency components. These observations motivated us to develop a Lo-Hi synergistic design framework that leverages both biases.

4. Lo-Hi Synergistic Design Framework

Our proposed synergistic design framework comprises two components: a Lo-Hi quantization process and a Lo-Hi compensation model. Based on our observations, the quantization process in traditional compression is contrary to the priority of the compensation network. Therefore, we first modify the quantization process to make it suitable for the compensation network. Then, we design a two-step compensation CNN framework with a proposed basis attention block (BAB) to separately address low and high FCs.

4.1. A Lo-Hi Sensitive Quantization Process

Quantization steps in existing image compression standards are derived from the non-linear mapping from a quantization parameter (QP) [6, 28, 34], or a default quantization table scaled by a quality factor (QF) [30] and QP [28]. The HVS-driven quantization table provides smaller steps for low FCs to adapt to HVS's high sensitivity in plain areas while giving larger steps for high frequency to discard dramatic changes. However, this setting contradicts the CNN compensation performance, as our observation suggests that low FCs are easier for CNNs to compensate.

To redesign the quantization process, we first analyze the distribution of each FC coefficient. The Generalized Gaussian Distribution with zero means is suitable for DCT coefficient modeling [43] with the distribution:

$$GG(x;\beta,\rho) = \frac{1/2}{2\Gamma(1+1/\rho)} e^{-\beta^{\rho/2}|x|^{\rho}}, \tag{2}$$

where ρ is the shape parameter and β is the inverse scale parameter. We follow the assumption as in [43], setting $\rho = 2$, to model the DCT coefficients as a Gaussian distribution. As shown in Fig. 3(b), the quantized symbols at low FCs are diverse after default compression since the quantization step (QS) is small. Our method replaces the default quantization step with the corresponding standard deviation (st.d) σ of this FC. In this way, the FCs with larger and smaller st.d will all have the same distribution after quantization, as shown in Fig. 3(c). Reducing the number of symbols at the encoding stage can decrease the compression bitrates.

We formulated the process of our quantization method and gave the ideal amount of entropy reduction benefit of our design as below. Taking the DCT as an example, the probability distribution of DCT coefficients can be generally denoted as:

$$P_i(x) = 1/\sqrt{2\pi}\sigma_i exp(-x^2/2\sigma_i^2) \sim N(0,\sigma_i^2), \tag{3}$$

where $i \in 1, 2, ...64$, are DCT FC indices. After being divided by the default quantization step q or standard deviation σ, the distribution is shown as Eq. (4) and Eq. (5):

$$P_i(\tilde{x}) = 1/\sqrt{2\pi}\tilde{\sigma}_i exp(-\tilde{x}^2/2\tilde{\sigma_i}^2) \sim N(0,\tilde{\sigma_i}^2), \tag{4}$$

$$P_i(\hat{x}) = 1/\sqrt{2\pi} exp(-\hat{x}^2/2) \sim N(0,\hat{\sigma_i}^2), \tag{5}$$

where $\tilde{x} = x/q$, $\hat{x} = x/\sigma$, $\tilde{\sigma_i} = \sigma_i/q$, and $\hat{\sigma_i} = 1$. These scaled coefficients are rounded into integers (quantization intervals t), which are further sent to the entropy coder. When the standard deviation σ is large, it is reasonable to assume that the probability distribution of quantization intervals remains the Gaussian function. Thus, the distribution of t obtained from the division of q follows $N(0,\tilde{\varsigma_i}^2)$, where the $\tilde{\varsigma_i}$ is related with $\tilde{\sigma_i}$. Similarly, the distribution of intervals quantified by σ follows $N(0,\hat{\varsigma_i}^2)$ with $\hat{\varsigma_i}$ related with $\hat{\sigma_i}$.

We use Shannon entropy [25] to measure the entropy reduction of coefficients benefited from our quantization normalization based on default quantization. In general, we consider Shannon entropy of one specific FC, $H(\tilde{t})$ for quantization by q and $H(\hat{t})$ for quantization by σ. The distributions of quantization intervals obtained from the division of q and σ follow $N(0,\tilde{\varsigma}^2)$ and $N(0,\hat{\varsigma}^2)$, respectively. Thus the ideal amount of entropy reduction can be given by:

$$H(\tilde{t})/H(\hat{t}) = 1/2\log_2(2\pi e) + \log_2(\tilde{\varsigma} - \hat{\varsigma}), \tag{6}$$

where standard deviations $\tilde{\varsigma}$ and $\hat{\varsigma}$ in second term are related to standard deviations of scaled coefficients before rounding

given by $\tilde{\sigma} = \frac{\sigma}{q}$, and $\hat{\sigma} = \frac{\sigma}{\sigma} = 1$, wherein σ is the standard deviation of DCT coefficients of the specific FC. The numerical difference between σ and q is often significant in low FCs, leading to a large amount of entropy reduction and a further higher compression ratio. The detailed formula derivation and analysis of the possible maximum range of entropy reduction can be referred to the supplementary.

We consider that at high-frequency bands, their standard deviations are always small and they usually have limited contribution to the overall image quality. Thus, to maximize the compression ratio we use st.d as the quantization step and keep the others unchanged, which can be written as:

$$Q = \begin{cases} \sigma & , \sigma > dQ \quad (7a) \\ dQ & , \text{otherwise} \quad (7b) \end{cases}$$

where Q is the new quantization step at a FC, σ is its st.d and dQ is the default quantizaiton step. Therefore, most of the replaced components lie on low frequency, which means that a higher compression rate can be achieved due to the loss of high-energy low-frequency information. We refer to all the replaced FCs as frequency of interest (FOI).

4.2. A Lo-Hi Compensation CNN Framework

4.2.1. Framework Overview

As shown in Fig. 4, our two-step Lo-Hi compensation framework consists of a Lo-regression model and a Hi-generative model, each tailored to leverage the strengths of different network types for optimal frequency compensation. The **Lo-regression model** comprises an algorithm-agnostic regression model and a basis attention block (BAB), targeting FOI compensation. Specifically, the lightweight BAB is proposed to prioritize low-frequency information that is intentionally dropped in our Lo-Hi compression. Notably, our framework does not impose restrictions on the type of regression model used, allowing compatibility with any image restoration regression model. The operation of the Lo-regression model is defined as follows:

$$I_{Base} = F_{BAB}(F_{re}(I_{in})), \quad (8)$$

where F_{re} and F_{BAB} represent the algorithm-agnostic regression model and BAB module, respectively.

The **Hi-generative model** is responsible for high-frequency compensation, enhancing fine details for improved perceptual quality. Since the low-frequency components are effectively compensated by the Lo-regression model, we employ only a lightweight generative model for high-frequency refinement. This component can be any generative model suited to high-frequency restoration. The reconstructed image is defined as:

$$I_{rec} = I_{Base} + F_{ge}(I_{Base}), \quad (9)$$

where F_{ge} denotes the algorithm-agnostic generative model.

4.2.2. Basis Attention Block (BAB).

The BAB is a lightweight CNN model capable of selectively compensating for FCs, allowing for more compensation of severely compressed FOIs in the Lo-Hi sensitive quantization. The BAB can be divided into three sub-blocks: FOI Filter, statistic frequency attention, and spatial restoration.

FOI Filter. We employ DCT basis functions of FOIs as frequency selectors to focus the compensation process specifically on these targeted lower frequencies. In particular, the output image from the regression model is multiplied by the FOI-specific DCT basis functions in a non-overlapping manner, effectively isolating and enhancing the FOI components. FOI filters enables the compensation network to selectively restore the most critical frequencies, generating FOI-aware features for the succeeding frequency attention module.

Statistic Frequency Attention. we design a statistical frequency attention mechanism to assign higher priority to frequency components (FCs) with greater quantization losses, as indicated by larger differences between the default quantization step (QS) and the modified QS used in Lo-Hi sensitive quantization. Specifically, we use the quantization step residual (QS residual) as a weighting factor, giving higher weights to FOIs with larger QS residuals to ensure these frequencies receive more attention during compensation. The QS residual R_i for i-th FOI is calculated as follows:

$$R_i = Norm(\sum_{dataset}(\sigma_i - dQ_i)), \quad i \in FOIs. \quad (10)$$

where σ_i and dQ_i are the modified QS and the default QS for the i-th FOI, respectively. We obtain the QS residuals and QS of FOIs for training from the statistical experiment on the LSDIR [18] dataset across JPEG, HEVC, and VVC. The detailed analysis and the statistical results can be referred to the supplementary.

Additionally, we incorporate the modified QS values by expanding each QS to match the feature resolution and concatenating them along the channel dimension. This setup enables the network to selectively focus on frequencies with larger QS values, corresponding to higher quantization losses. Finally, we downsample the output three times by a factor of 2 to align with the original image size, ready for subsequent processing.

Spatial Restoration. The Spatial Restoration module integrates the frequency-compensated features into the image reconstructed from the regression model, addressing the limitations of previous regression models in compensating FOIs. First, the output from the last module is concatenated with I_{Lo} along the channel dimension. Then, the AOT block [39], with its large receptive field, helps recover information lost during downsampling. Next, the spatial attention module

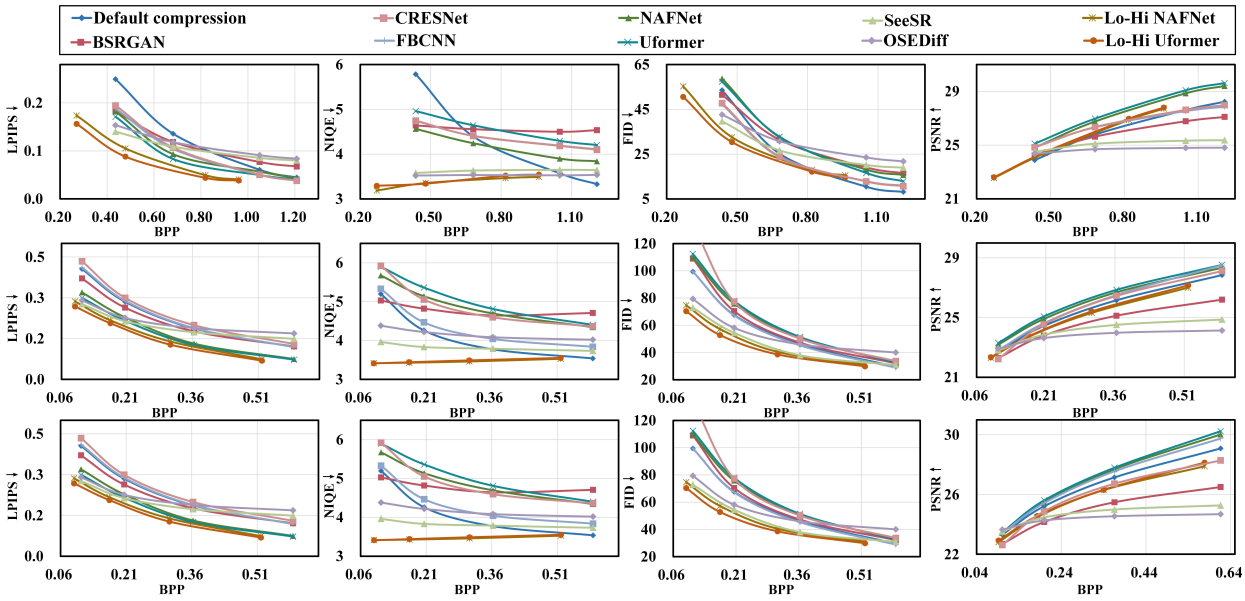


Figure 5. Rate distortion curves under JPEG, HEVC, and VVC compression (from top to bottom, respectively). LPIPS, NIQE, FID, and PSNR are adopted to measure the compensation quality (from left to right, respectively). The corresponding BD rates are reported in Table 1.

extracts spatial information to complement the frequency-focused compensation from the statistical frequency attention. Finally, these enhanced features are fused through a convolution layer to produce I_{Base}.

5. Experiment

5.1. Experimental Setup

Our synergistic design framework can be applied to any traditional compression method. In this paper, we implement the framework in JPEG, HEVC, and VVC. We replace the default quantization process of the Y channel in these codecs with our Lo-Hi sensitive quantization.

Dataset and metrics. The experiments were conducted on LSDIR dataset [18], including 50k images as the training dataset and 250 as the testing dataset. Two datasets are generated, one using the default quantization and the other using our Lo-Hi sensitive quantization. For JPEG, the QFs of 10, 20, 40, and 50 are considered, and for HEVC and VVC, the QPs of 33, 37, 41, and 45 are considered. BD-rate is used as the compression performance metric. More negative BD-rate corresponds to higher average bitrate savings for the same visual quality. We adopt PSNR, NIQE, LPIPS, and FID to measure human vision quality and mean average precision (mAP) for the object detection task. The baseline encoder of all the BD-rate calculations in our experiments is the default compression of the respective compression standard.

Models and Training. To verify the generality of our framework, we evaluate the performance with different models. For the heavy-weight regression model, we choose the recent SOTA models, Uformer [33] and NAFNet [9]. For the light-weight Hi-generative model, we only use 12 RRDB blocks to form a simple model [41] as the generator and UNet with spectral normalization as the discriminator [41]. We simply refer to them as Lo-Hi NAFNet and Lo-Hi Uformer. We train these models using NVIDIA A6000 GPUs, setting a total batch size of 8 and a patch size of 256×256. We first train the Lo-regression model with loss function of L_2. Then with Lo-regression model frozen, we train the Hi-generative model with loss function of $L_1 + 0.1 * L_{GAN} + 0.1 * L_{perc}$. The training iterations of regression models are 2×10^5 and 5×10^4 for the generative model. The learning rate is 1×10^{-4}. We evaluate seven SOTA compensation networks as baselines, including regression models: NAFNet [9], Uformer [33], FBCNN [15], and CRESNet [11], and generative models: GAN-based BSRGAN [41], diffusion-based SeeSR [36], and diffusion-based OSEDiff [35].

5.2. Ablation Study

5.2.1. Lo-Hi Sensitive Quantization.

We use BD-rate to evaluate compression performance in terms of bitrates and quality, and calculate pure bitrate savings (PBS) averaged across 4 QPs/QFs, to assess bitrate savings independently of quality. Comparison results between our Lo-Hi sensitive compression and the defualt compression are presented in Table 2, with corresponding rate distortion (RD) curves provided in supplementary.

Our Lo-Hi compression effectively mitigates FOIs, achieving up to 27.55% PBS compared to default JPEG compression. Additionally, it consistently yields BD-rate reduction over LPIPS across the three compression standards,

Table 1. BD rates and model sizes of compression and compensation frameworks. More negative BD-rate means higher savings.

Compression	Compensation	Size(M)	BD-rate (JPEG)↓			BD-rate (HEVC)↓			BD-rate (VVC)↓		
			LPIPS	FID	PSNR	LPIPS	FID	PSNR	LPIPS	FID	PSNR
Default	BSRGAN [41]	16.8	-7.29%	5.54%	5.56%	-9.80%	7.64%	34.52%	-3.74%	6.37%	58.94%
Default	SeeSR [36]	2756.3	-10.67%	-1.05%	7.61%	-29.09%	-26.61%	31.41%	-22.82%	-14.61%	35.19%
Default	OSEDiff [35]	2392.1	-6.38%	2.99%	13.22%	-24.14%	-16.00%	35.38%	-16.35%	-10.91%	37.83%
Default	FBCNN [15]	71.9	-14.41%	-4.43%	-10.61%	1.54%	-1.01%	-7.97%	1.09%	-0.55%	-8.63%
Default	CRESNet [11]	55.9	-13.40%	-4.55%	-10.95%	9.02%	21.12%	-3.56%	21.33%	21.07%	14.94%
Default	NAFNet [9]	67.8	-23.29%	12.65%	-21.02%	-30.21%	17.87%	-15.07%	-28.22%	3.71%	-10.90%
Default	Uformer [33]	50.4	-26.54%	9.22%	**-25.02%**	-34.99%	21.20%	**-19.44%**	-35.86%	7.38%	**-13.47%**
Lo-Hi	Lo-Hi NAFNet	76.5	-40.09%	-23.09%	-5.301%	-43.77%	-33.31%	10.49%	-42.09%	-32.41%	19.79%
Lo-Hi	Lo-Hi Uformer	59.1	**-43.68%**	**-27.55%**	-6.930%	**-47.91%**	**-39.53%**	7.62%	**-45.62%**	**-40.04%**	17.00%

Table 2. BD rates and pure bitrate savings (PBS) of the Lo-Hi sensitive compression compared with default compression. For **BD rates** over LPIPS and PSNR, more **negative** indicates more savings. For **PBS**, more **positive** means more savings.

JPEG			HEVC			VVC		
PSNR	LPIPS	PBS	PSNR	LPIPS	PBS	PSNR	LPIPS	PBS
-1.27%	-11.50%	27.55%	2.51%	-6.79%	14.65%	2.12%	-1.56%	7.45%

Table 3. Ablation study of Lo-Hi Uformer framework

Regression	BAB	Generative	JPEG	
50.39 M	0.05 M	8.71 M	LPIPS	PSNR
✘	✘	✔	-5.21%	52.31%
✔	✘	✘	-26.54%	-25.02%
✔	✔	✘	-28.23%	-49.43%
✔	✔	✔	-43.68%	-6.93%

Since BAB is a block specifically designed to compensate for Lo-Hi Compression, the use of BAB requires pairing with Lo-Hi compression.

with the maximum BD-rate increase of only 2.51% over PSNR, indicating minimal impact on perceptual quality.

Notably, as compression standards advance, both PBS and BD-rate savings from our Lo-Hi compression decrease, reflecting the expected trade-off between computational complexity and compression ratio. The additional compression latency from replacing QS is negligible, adding only 0.003s for JPEG, as shown in Fig. 1. This additional latency proportion is significantly reduced in HEVC and VVC, due to their higher baseline computational complexity.

5.2.2. Compensation network

We conduct ablation experiments of the Lo-Hi Uformer compensation framework on JPEG compression, with results shown in Table 3. As can be seen, BD-rate performance on LPIPS and PSNR is inferior when using only the lightweight generative model. Compared to the regression model alone, adding BAB to the regression model yields a 1.69% improvement in BD-rate over LPIPS and a substantial 24.41% improvement over PSNR, with only an additional 0.05M parameters. This demonstrates that BAB effectively enhances image quality through FOI compensation, particularly in terms of fidelity. Additionally, the full Lo-Hi framework, which includes the BAB and the generative model, achieves a 17.14% improvement in BD-rate over LPIPS compared to the regression model alone, while adding only 8.76M parameters. Although the generative model alone places less emphasis on fidelity, combining it with the regression model and BAB results in a significant 59.24% improvement in BD-rate over PSNR compared to the generative model alone, highlighting the importance of synergistic integration for both perceptual quality and fidelity.



Figure 6. Qualitative comparisons of different generative models. Three Hi models are evaluated on images that are output from our Lo-regression model. Our GAN-based Hi-generative model exhibits faithful, clear, and visually pleasing results.

We additionally evaluated the perceptual enhancement performance of different generative models serving as the Hi-generative model in our framework. We evaluate diffusion-based SeeSR and OSEDiff on the images reconstructed by our Lo-regression model. As seen in Fig. 6, though input images have been reconstructed by the Lo-regression model, diffusion-based models still struggle with rendering accurate and coherent text, aligning with common findings in diffusion models [8, 44]. Instead, our lightweight GAN model delivers more faithful and clear visual results. Future efforts will be involved in including reliable diffusion models in our Lo-Hi framework.

5.3. Overall Performance Comparison

In the comparison of overall frameworks, the Lo-Hi compensation frameworks restore images compressed by our Lo-Hi sensitive compression while baseline models restore images by default compression. The RD curves and BD-rate performance are presented in Fig. 5 and Table 1, respectively. Compared to default compression, our Lo-Hi frameworks achieve significant BD-rate reduction up to **47.91%** (Lo-Hi Uformer on HEVC) over LPIPS and **40.04%** (Lo-Hi Uformer on VVC) over FID, while other evaluated models offer BD-rate reduction no more than 35.86% (Uformer on VVC) over LPIPS and 26.61% over FID (SeeSR on HEVC). Besides, our frameworks deliver the fabulously lowest NIQE values as shown in Fig. 5. Fig. 7 illustrates the qualitative comparisons, from which we can see that our Lo-Hi synergistic design framework offers high perceptual quality in the context of reality and details, though the rate-distortion per-

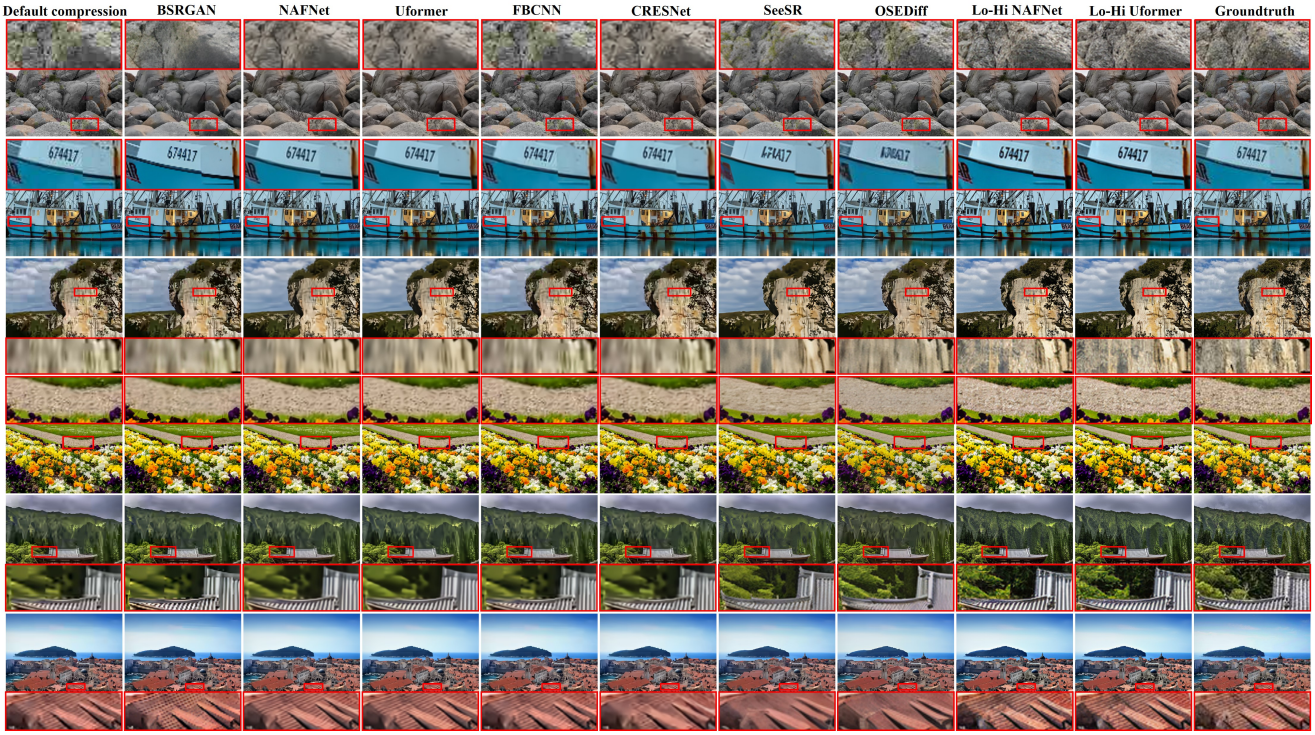


Figure 7. Qualitative comparisons of compensation frameworks under different compression conditions. Images compressed with JPEG at $QF = 10$, JPEG at $QF = 40$, HEVC at $QP = 45$, HEVC at $QP = 37$, VVC at $QP = 45$, VVC at $QP = 37$, are illustrated from top to bottom. Our Lo-Hi synergistic design frameworks produce more clear, detailed, and faithful results. Zoom in for better visual comparison.

Table 4. BD rates over **mAP** on the COCO dataset.

Codec	BSRGAN	NAFNet	Lo-Hi NAFNet
JPEG	-29.96%	-31.22%	**-48.12%**
HEVC	19.40%	-24.54%	**-41.63%**
VVC	8.77%	-6.32%	**-54.53%**

formance measured by PSNR is barely satisfactory shown in Table 1. In other words, the remarkable perceptual quality benefited from our framework comes at the cost of increased pixel distortion, which is aligned with the well-known rate-distortion-perception trade-off [4, 5, 40, 45]. More visual results are available in supplementary.

In summary, our Lo-Hi sensitive quantization enables substantially reduced bitrates with acceptable quality impact, resulting in efficient compression. The Lo-regression model in our synergistic design effectively compensates for the deliberately quantized low-frequency components, enhancing fidelity by restoring critical information. Meanwhile, the Hi-generative model further elevates perceptual quality, allowing our framework to achieve a balanced optimization across fidelity and perceptual metrics.

5.4. Machine Vision Task

Our Lo-Hi synergistic design framework excels not only in the human visual domain [19] but also demonstrates outstanding performance in object detection tasks. We conducted experiments on a set of 100 images from the COCO dataset using DAMO-YOLO [37] for object detection. As seen in Table 4, our Lo-Hi NAFNet achieves consistent and promising BD-rate savings over mean average precision (mAP), up to **54.53%** BD-rate reduction in VVC. Notably, our Lo-Hi framework exhibits a greater improvement in BD-rate on VVC than JPEG and HEVC, which we attribute to VVC's ability to retain more information at same BPP levels. Thus our framework can leverage the frequency bias and the increased retained information in VVC to enhance detection results. Visual examples are available in supplementary.

6. Conclusion

We identify a frequency bias mismatch between image compression and compensation, revealing low-frequency redundancies in existing CAR frameworks. These insights led us to develop a frequency-aware Lo-Hi synergistic framework. Our approach includes a Lo-Hi sensitive quantization process to compress more low FCs and a two-step Lo-Hi compensation framework to address low and high FCs separately. Moreover, we propose a basis attention block to prioritize intentionally dropped low-frequency information. Experimental results demonstrate that our framework effectively combines the strengths of conventional codecs with CNNs' unique properties, achieving superior compression and compensation performance.

References

[1] Johannes Ballé, Valero Laparra, and Eero P Simoncelli. End-to-end optimized image compression. *arXiv preprint arXiv:1611.01704*, 2016.

[2] Johannes Ballé, David Minnen, Saurabh Singh, Sung Jin Hwang, and Nick Johnston. Variational image compression with a scale hyperprior. *arXiv preprint arXiv:1802.01436*, 2018.

[3] Kashif Bilal, Osman Khalid, Aiman Erbad, and Samee U Khan. Potentials, trends, and prospects in edge technologies: Fog, cloudlet, mobile edge, and micro data centers. *Computer Networks*, 130:94–120, 2018.

[4] Yochai Blau and Tomer Michaeli. The perception-distortion tradeoff. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6228–6237, 2018.

[5] Yochai Blau and Tomer Michaeli. Rethinking lossy compression: The rate-distortion-perception tradeoff. In *International Conference on Machine Learning*, pages 675–685, 2019.

[6] Benjamin Bross, Ye-Kui Wang, Yan Ye, Shan Liu, Jianle Chen, Gary J Sullivan, and Jens-Rainer Ohm. Overview of the versatile video coding (VVC) standard and its applications. *IEEE Transactions on Circuits and Systems for Video Technology*, 31(10):3736–3764, 2021.

[7] Keshigeyan Chandrasegaran, Ngoc-Trung Tran, and Ngai-Man Cheung. A closer look at fourier spectrum discrepancies for CNN-generated images detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7200–7209, 2021.

[8] Jingye Chen, Yupan Huang, Tengchao Lv, Lei Cui, Qifeng Chen, and Furu Wei. TextDiffuser: Diffusion models as text painters. In *Advances in Neural Information Processing Systems*, pages 9353–9387. Curran Associates, Inc., 2023.

[9] Liangyu Chen, Xiaojie Chu, Xiangyu Zhang, and Jian Sun. Simple baselines for image restoration. In *European Conference on Computer Vision*, pages 17–33, 2022.

[10] Tong Chen, Haojie Liu, Zhan Ma, Qiu Shen, Xun Cao, and Yao Wang. End-to-end learnt image compression via non-local attention optimization and improved context modeling. *IEEE Transactions on Image Processing*, 30:3179–3191, 2021.

[11] Yiyun Chen, Yunmeng Liu, Mingliang Chen, Zirui Wang, Wenming Yang, and Qingmin Liao. Blind JPEG compression artifacts removal by integrating channel regulation with exit strategy. *IEEE Transactions on Multimedia*, 2022.

[12] Zhijian Hao, Heming Sun, Guohao Xu, Jiaming Liu, Xiankui Xiong, Xuanpeng Zhu, Xiaoyang Zeng, and Yibo Fan. Fast transform kernel selection based on frequency matching and probability model for av1. *IEEE Transactions on Broadcasting*, 2024.

[13] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in neural information processing systems*, 30, 2017.

[14] Jiahuan Ji, Baojiang Zhong, Weigang Song, and Kai-Kuang Ma. Learning multi-scale features for JPEG image artifacts removal. In *2023 IEEE International Conference on Image Processing (ICIP)*, pages 1565–1569, 2023.

[15] Jiaxi Jiang, Kai Zhang, and Radu Timofte. Towards flexible blind JPEG artifacts removal. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4997–5006, 2021.

[16] Steffen Jung and Margret Keuper. Spectral distribution aware image generation. In *Proceedings of the AAAI conference on artificial intelligence*, pages 1734–1742, 2021.

[17] Fengyong Li, Huajun Zhai, Teng Liu, Xinpeng Zhang, and Chuan Qin. Learning compressed artifact for JPEG manipulation localization using wide-receptive-field network. *ACM Transactions on Multimedia Computing, Communications and Applications*, 2024.

[18] Yawei Li, Kai Zhang, Jingyun Liang, Jiezhang Cao, Ce Liu, Rui Gong, Yulun Zhang, Hao Tang, Yun Liu, Denis Demandolx, et al. Lsdir: A large scale dataset for image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1775–1787, 2023.

[19] Jiaming Liu, Zihao Liu, Xuan Huang, Ruoxi Zhu, Qi Zheng, Zhijian Hao, Tao Liu, Jun Tao, and Yibo Fan. Auto-isp: An efficient real-time automatic hyperparameter optimization framework for isp hardware system. In *Proceedings of the 61st ACM/IEEE Design Automation Conference*, pages 1–6, 2024.

[20] Anish Mittal, Rajiv Soundararajan, and Alan C. Bovik. Making a "completely blind" image quality analyzer. *IEEE Signal Processing Letters*, 20(3):209–212, 2013.

[21] Maxime Oquab, Pierre Stock, Daniel Haziza, Tao Xu, Peizhao Zhang, Onur Celebi, Yana Hasson, Patrick Labatut, Bobo Bose-Kolanu, Thibault Peyronel, and Camille Couprie. Low bandwidth video-chat compression using deep generative models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 2388–2397, 2021.

[22] Nasim Rahaman, Aristide Baratin, Devansh Arpit, Felix Draxler, Min Lin, Fred Hamprecht, Yoshua Bengio, and Aaron Courville. On the spectral bias of neural networks. In *Proceedings of the 36th International Conference on Machine Learning*, pages 5301–5310, 2019.

[23] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. ImageNet large scale visual recognition challenge. *International Journal of Computer Vision (IJCV)*, 115 (3):211–252, 2015.

[24] Katja Schwarz, Yiyi Liao, and Andreas Geiger. On the frequency bias of generative models. *Advances in Neural Information Processing Systems*, 34:18126–18136, 2021.

[25] Claude Elwood Shannon. A mathematical theory of communication. *ACM SIGMOBILE mobile computing and communications review*, 5(1):3–55, 2001.

[26] Zhanjun Si and Ke Shen. Research on the WebP image format. In *Advanced graphic communications, packaging technology and materials*, pages 271–277, 2016.

[27] Athanassios N Skodras, Charilaos A Christopoulos, and Touradj Ebrahimi. JPEG2000: The upcoming still image compression standard. *Pattern recognition letters*, 22(12): 1337–1345, 2001.

[28] Gary J Sullivan, Jens-Rainer Ohm, Woo-Jin Han, and Thomas Wiegand. Overview of the high efficiency video coding (HEVC) standard. *IEEE Transactions on circuits and systems for video technology*, 22(12):1649–1668, 2012.

[29] Lucas Theis, Wenzhe Shi, Andrew Cunningham, and Ferenc Huszár. Lossy image compression with compressive autoencoders. *arXiv preprint arXiv:1703.00395*, 2017.

[30] Gregory K Wallace. The JPEG still picture compression standard. *Communications of the ACM*, 34(4):30–44, 1991.

[31] Xintao Wang, Liangbin Xie, Chao Dong, and Ying Shan. Real-ESRGAN: Training real-world blind super-resolution with pure synthetic data. In *IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, pages 1905–1914, 2021.

[32] Xi Wang, Xueyang Fu, Yurui Zhu, and Zheng-Jun Zha. JPEG artifacts removal via contrastive representation learning. In *European Conference on Computer Vision*, pages 615–631, 2022.

[33] Zhendong Wang, Xiaodong Cun, Jianmin Bao, Wengang Zhou, Jianzhuang Liu, and Houqiang Li. Uformer: A general u-shaped transformer for image restoration. In *Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition*, pages 17683–17693, 2022.

[34] Thomas Wiegand, Gary J Sullivan, Gisle Bjontegaard, and Ajay Luthra. Overview of the h. 264/AVC video coding standard. *IEEE Transactions on Circuits and Systems for Video Technology*, 13(7):560–576, 2003.

[35] Rongyuan Wu, Lingchen Sun, Zhiyuan Ma, and Lei Zhang. One-step effective diffusion network for real-world image super-resolution. *arXiv preprint arXiv:2406.08177*, 2024.

[36] Rongyuan Wu, Tao Yang, Lingchen Sun, Zhengqiang Zhang, Shuai Li, and Lei Zhang. Seesr: Towards semantics-aware real-world image super-resolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 25456–25467, 2024.

[37] Xianzhe Xu, Yiqi Jiang, Weihua Chen, Yilun Huang, Yuan Zhang, and Xiuyu Sun. Damo-yolo: A report on real-time object detection design. *arXiv preprint arXiv:2211.15444*, 2022.

[38] Juncheol Ye, Hyunho Yeo, Jinwoo Park, and Dongsu Han. Accelir: Task-aware image compression for accelerating neural restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18216–18226, 2023.

[39] Yanhong Zeng, Jianlong Fu, Hongyang Chao, and Baining Guo. Aggregated contextual transformations for high-resolution image inpainting. *IEEE Transactions on Visualization and Computer Graphics*, pages 1–1, 2022.

[40] George Zhang, Jingjing Qian, Jun Chen, and Ashish Khisti. Universal rate-distortion-perception representations for lossy compression. *Advances in Neural Information Processing Systems*, 34:11517–11529, 2021.

[41] Kai Zhang, Jingyun Liang, Luc Van Gool, and Radu Timofte. Designing a practical degradation model for deep blind image super-resolution. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4791–4800, 2021.

[42] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 586–595, 2018.

[43] Xinfeng Zhang, Weisi Lin, Ruiqin Xiong, Xianming Liu, Siwei Ma, and Wen Gao. Low-rank decomposition-based restoration of compressed images via adaptive noise estimation. *IEEE Transactions on Image processing*, 25(9):4158–4171, 2016.

[44] Yiming Zhao and Zhouhui Lian. UDiffText: A unified framework for high-quality text synthesis in arbitrary images via character-aware diffusion models. In *Computer Vision – ECCV 2024*, pages 217–233, Cham, 2025. Springer Nature Switzerland.

[45] Qi Zheng, Yibo Fan, Leilei Huang, Tianyu Zhu, Jiaming Liu, Zhijian Hao, Shuo Xing, Chia-Ju Chen, Xiongkuo Min, Alan C. Bovik, and Zhengzhong Tu. Video quality assessment: A comprehensive survey, 2024.

[46] Qi Zheng, Haozhi Wang, Zihao Liu, Jiaming Liu, Peiye Liu, Zhijian Hao, Yanheng Lu, Dimin Niu, Jinjia Zhou, Minge Jing, and Yibo Fan. Unicorn: Unified neural image compression with one number reconstruction, 2024.