Frequency-Biased Synergistic Design for Image Compression and Compensation

Supplementary Material

In this supplementary material, we first provide solid formula proof of the effectiveness of Lo-Hi sensitive quantization. Then we introduce the statistical information on frequency of interest (FOI) and how they contribute to our basis attention block (BAB). Finally, we present more comprehensive experimental results, including rate-distortion curves of evaluated compression algorithms, qualitative results of compensation frameworks, rate-mAP curves and visual results of object detection, and complexity analysis of BAB.

1. Formula Proof of Effectiveness of Lo-Hi Sensitive Quantization

In general, we consider one specific frequency component (FC) of the DCT transform. Hence its intervals quantified by default step q and standard deviation of this FC σ are denoted as $\tilde{t}$ and $\hat{t}$, approximately following $N(0,\tilde{\varsigma})$ and $N(0,\hat{\varsigma})$ [3], respectively. We use Shannon entropy [2] to measure the entropy reduction of coefficients benefitted from our Lo-Hi sensitive quantization in comparison with default quantization. Specifically, the entropy of intervals obtained from default quantization can be given by:

$$H(\tilde{t}) = -\sum P(\tilde{t})\log_2 P(\tilde{t}) \tag{1}$$

$$= -\sum_{n=1}^{N} \frac{1}{\sqrt{2\pi}\tilde{\varsigma}} exp(-\frac{\tilde{t}_n^2}{2\tilde{\varsigma}^2}) \log_2 \frac{1}{\sqrt{2\pi}\tilde{\varsigma}} exp(-\frac{\tilde{t}_n^2}{2\tilde{\varsigma}^2}), \tag{2}$$

where N is the number of all possible values of the quantization intervals within the overall image. Correspondingly, the entropy of intervals derived from Lo-Hi sensitive quantization is denoted as:

$$H(\hat{t}) = -\sum_{m=1}^{M} \frac{1}{\sqrt{2\pi}\hat{\varsigma}} exp(-\frac{\hat{t}_m^2}{2\hat{\varsigma}^2}) \log_2 \frac{1}{\sqrt{2\pi}\hat{\varsigma}} exp(-\frac{\hat{t}_m^2}{2\hat{\varsigma}^2}), \tag{3}$$

and M is the number of all possible values of intervals. We measure the entropy reduction as follows:

$$\frac{H(\tilde{t})}{H(\hat{t})} = \frac{-\sum_{n=1}^{N} \frac{1}{\sqrt{2\pi}\tilde{\varsigma}} exp(-\frac{\tilde{t}_n^2}{2\tilde{\varsigma}^2}) \log_2 \frac{1}{\sqrt{2\pi}\tilde{\varsigma}} exp(-\frac{\tilde{t}_n^2}{2\tilde{\varsigma}^2})}{-\sum_{m=1}^{M} \frac{1}{\sqrt{2\pi}\hat{\varsigma}} exp(-\frac{\hat{t}_m^2}{2\hat{\varsigma}^2}) \log_2 \frac{1}{\sqrt{2\pi}\hat{\varsigma}} exp(-\frac{\hat{t}_m^2}{2\hat{\varsigma}^2})}. \tag{4}$$

To reduce the Equation 4, we first simplify Equation 2:

$$H(\tilde{t}) = -\sum_{n=1}^{N} \frac{1}{\sqrt{2\pi}\tilde{\varsigma}} exp(-\frac{\tilde{t}_n^2}{2\tilde{\varsigma}^2}) \log_2 \frac{1}{\sqrt{2\pi}\tilde{\varsigma}} exp(-\frac{\tilde{t}_n^2}{2\tilde{\varsigma}^2}) \tag{5}$$

$$= -\sum_{n=1}^{N} \frac{1}{\sqrt{2\pi}\tilde{\varsigma}} exp(-\frac{\tilde{t}_n^2}{2\tilde{\varsigma}^2})(-\log_2(\sqrt{2\pi}\tilde{\varsigma}) - \frac{\tilde{t}_n^2}{2\tilde{\varsigma}^2}\log_2 e) \tag{6}$$

$$= \log_2(\sqrt{2\pi}\tilde{\varsigma}) \sum_{n=1}^{N} \frac{1}{\sqrt{2\pi}\tilde{\varsigma}} exp(-\frac{\tilde{t}_n^2}{2\tilde{\varsigma}^2}) \tag{7}$$

$$+ \log_2 e \sum_{n=1}^{N} \frac{1}{\sqrt{2\pi}\tilde{\varsigma}} exp(-\frac{\tilde{t}_n^2}{2\tilde{\varsigma}^2}) \frac{\tilde{t}_n^2}{2\tilde{\varsigma}^2}. \tag{8}$$

Note that the summation of the first term in Equation 8 sums the probability of $\{\tilde{t}\}_n^N$:

$$\sum_{n=1}^{N} \frac{1}{\sqrt{2\pi}\tilde{\varsigma}} exp(-\frac{\tilde{t}_n^2}{2\tilde{\varsigma}^2}) = 1. \tag{9}$$

Therefore, Equation 8 can be further simplified as:

$$H(\tilde{t}) = \log_2(\sqrt{2\pi}\tilde{\varsigma}) \sum_{n=1}^{N} \frac{1}{\sqrt{2\pi}\tilde{\varsigma}} exp(-\frac{\tilde{t}_n^2}{2\tilde{\varsigma}^2}) \tag{10}$$

$$+ \log_2 e \sum_{n=1}^{N} \frac{1}{\sqrt{2\pi}\tilde{\varsigma}} exp(-\frac{\tilde{t}_n^2}{2\tilde{\varsigma}^2}) \frac{\tilde{t}_n^2}{2\tilde{\varsigma}^2} \tag{11}$$

$$= \log_2(\sqrt{2\pi}\tilde{\varsigma}) + \log_2 e \sum_{n=1}^{N} \frac{1}{\sqrt{2\pi}\tilde{\varsigma}} exp(-\frac{\tilde{t}_n^2}{2\tilde{\varsigma}^2}) \frac{\tilde{t}_n^2}{2\tilde{\varsigma}^2} \tag{12}$$

$$= \log_2(\sqrt{2\pi}\tilde{\varsigma}) + \frac{\log_2 e}{2\tilde{\varsigma}^2} \sum_{n=1}^{N} \frac{1}{\sqrt{2\pi}\tilde{\varsigma}} exp(-\frac{\tilde{t}_n^2}{2\tilde{\varsigma}^2}) \tilde{t}_n^2. \tag{13}$$

Obviously, the summation of the second term in Equation 13 calculates the variance of $\{\tilde{t}\}_n^N$:

$$\sum_{n=1}^{N} \frac{1}{\sqrt{2\pi}\tilde{\varsigma}} exp(-\frac{\tilde{t}_n^2}{2\tilde{\varsigma}^2}) \tilde{t}_n^2 = \tilde{\varsigma}^2, \tag{14}$$

with which Equation 13 is further reduced as:

$$H(\tilde{t}) = \log_2(\sqrt{2\pi}\tilde{\varsigma}) + \frac{\log_2 e}{2\tilde{\varsigma}^2}\sum_{n=1}^{N}\frac{1}{\sqrt{2\pi}\tilde{\varsigma}}exp(-\frac{\tilde{t}_n^2}{2\tilde{\varsigma}^2})\tilde{t}_n^2 \tag{15}$$

$$= \log_2(\sqrt{2\pi}\tilde{\varsigma}) + \frac{\log_2 e}{2\tilde{\varsigma}^2}\tilde{\varsigma}^2 \tag{16}$$

$$= \log_2(\sqrt{2\pi}\tilde{\varsigma}) + \frac{\log_2 e}{2} \tag{17}$$

$$= \frac{1}{2}\log_2(2\pi e) + \log_2\tilde{\varsigma}. \tag{18}$$

Similarly, Equation 3 can be simplified into:

$$H(\hat{t}) = \frac{1}{2}\log_2(2\pi e) + \log_2\hat{\varsigma}. \tag{19}$$

With Equation 18 and Equation 19, the entropty reduction denoted by Equation 4 can be reduced as:

$$\frac{H(\tilde{t})}{H(\hat{t})} = \frac{-\sum_{n=1}^{N}\frac{1}{\sqrt{2\pi}\tilde{\varsigma}}exp(-\frac{\tilde{t}_n^2}{2\tilde{\varsigma}^2})\log_2\frac{1}{\sqrt{2\pi}\tilde{\varsigma}}exp(-\frac{\tilde{t}_n^2}{2\tilde{\varsigma}^2})}{-\sum_{m=1}^{M}\frac{1}{\sqrt{2\pi}\hat{\varsigma}}exp(-\frac{\hat{t}_m^2}{2\hat{\varsigma}^2})\log_2\frac{1}{\sqrt{2\pi}\hat{\varsigma}}exp(-\frac{\hat{t}_m^2}{2\hat{\varsigma}^2})} \tag{20}$$

$$= \frac{\frac{1}{2}\log_2(2\pi e) + \log_2\tilde{\varsigma}}{\frac{1}{2}\log_2(2\pi e) + \log_2\hat{\varsigma}} \tag{21}$$

$$= \frac{1}{2}\log_2(2\pi e) + \log_2(\tilde{\varsigma} - \hat{\varsigma}). \tag{22}$$

Equation 22 gives the ideal reduction amount of entropy induced by our Lo-Hi sensitive quantization. The first term approximately equals 2.05. The standard deviations $\tilde{\varsigma}$ and $\hat{\varsigma}$ in the second term are related to standard deviations of scaled coefficients before rounding given by $\tilde{\sigma} = \frac{\sigma}{q}$ and $\tilde{\sigma} = \frac{\sigma}{\sigma} = 1$, wherein σ is the standard deviation of DCT coefficients of the specific FC. Note that the replacement of quantization step happens only in low frequency components according to our statistical results. Moreover, The numerical difference between q and σ is often large in low frequency components, leading to the large amount of entropy reduction and further higher compression ratio.

By the numbers, we analyze the possible maximum range of entropy reduction. Specifically, we follow the JPEG compression process at $QF = 60$ and take a random image from the DIV2K [1] dataset as an instance. We obtain two sets of integer values from default quantization and statistic-based quantization, respectively. The minimum amount of entropy reduction is obtained as $\frac{1}{2}\log_2(2\pi e) \approx 2.05$ where $Q = dQ$. And the maximum amount of entropy reduction is obtained at the first AC component where $Q = \delta$. The $\tilde{\varsigma}$ and $\hat{\varsigma}$ in Equation.4 can be estimated from the above integer values, thus the amount of entropy reduction is around 4.65. Therefore, the range of ideal entropy reduction benefitted from our Lo-Hi sensitive quantization normalization is $[2.05, 4.65]$.

2. Statistical Information of Frequency of Interest (FOIs)

In this section, we describe how we collect statistical information of FOIs applied to the basis attention block (BAB).

In Lo-Hi sensitive quantization process, we replace the default quantization step dQ with the standard deviation δ for each frequency component following:

$$Q = \begin{cases} \delta & , \delta > dQ \quad (23a) \\ dQ & , \text{otherwise}, \quad (23b) \end{cases}$$

which results in more information loss on low-frequency bands and a higher compression ratio. In the proposed compensation framework, we make the best of the quantization steps and delicately design the basis attention block accounting for recovering frequency information.

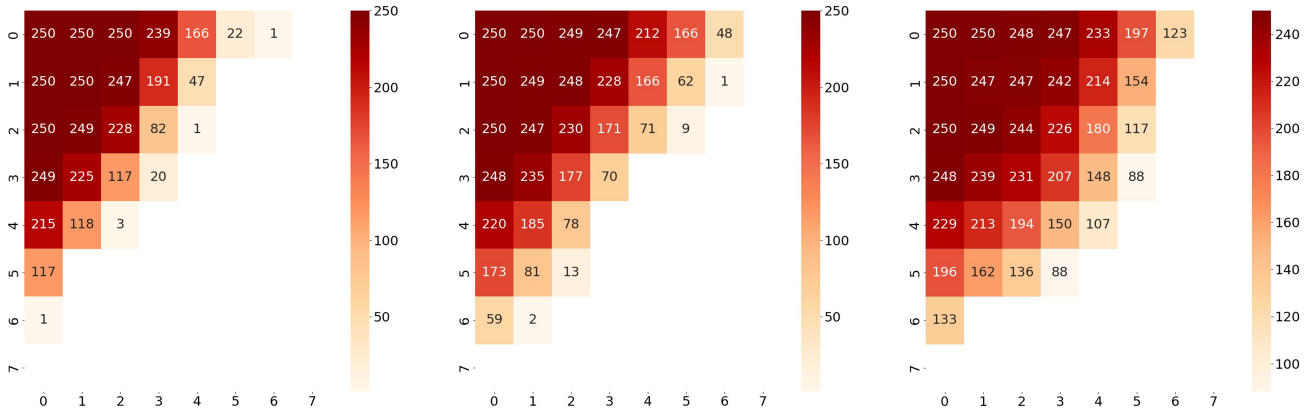
We conduct the JPEG, HEVC, and VVC compression on the testing dataset consisting of 250 images. The standard deviation of each frequency component is calculated and compared with the default quantization step (QS) of this FC. Across the datasets, the times when the replacement of the quantization step of each FC happens are counted, as shown in Fig. 1(a). Furthermore, the residual of the quantization step after the replacement is summed and normalized in a range of $[0, 1]$, as shown in Fig. 1(b). The residual is calculated as follows:

$$R_i = Norm(\sum_{dataset}(\delta_i - dQ_i)), \quad i \in FOIs. \tag{24}$$

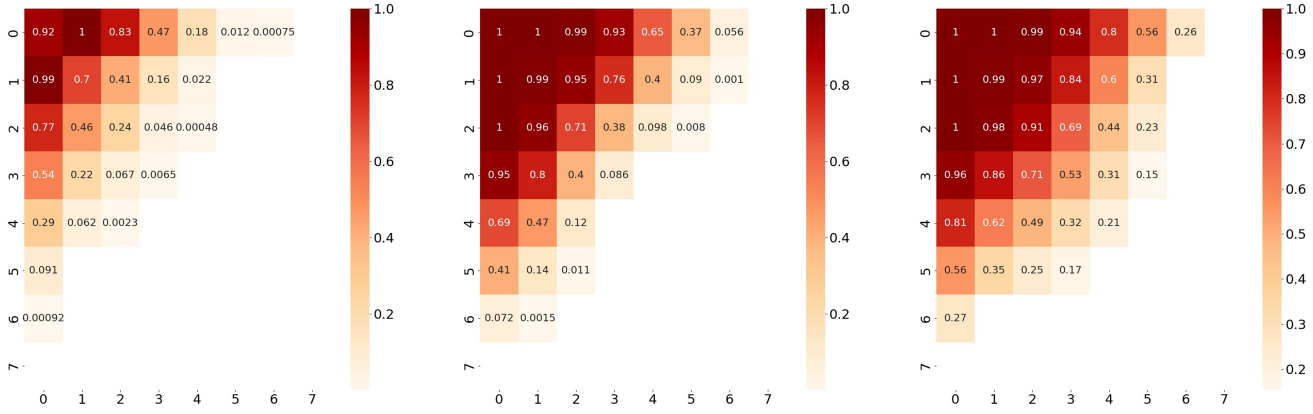
Note that the blank FCs in Fig. 1 indicate that no replacement of the quantization steps ever happens.

DCT basis function in BAB. It can be easily seen from Fig. 1 (a) that quantization steps are modified mainly in AC low-frequency components, which we named Frequency of Interest (FOI) in the paper. In the Lo-Hi compression method, JPEG compression has 26 FOIs, HEVC compression has 31 FOIs, and VVC compression has 34 FOIs Therefore, we employ these DCT basis functions in the basis attention block to extract the FOIs of an image, and focus more on recovering information within these FCs.

QS residual in BAB. In the Lo-Hi compensation, we also pay attention to the overall variations of quantization steps. The residuals of standard deviation and default quantization step indicate which FCs are lost more in the FOIs. Therefore, we use the normalization residuals in Fig. 1 (b) as the weights of these FOIs to strongly compensate for the frequency with a larger residual.



(a) The index and modification times of 8×8 DCT QS



(b) Normalization residual of 8×8 DCT QS

Figure 1. The statistical information of FOIs in our Lo-Hi sensitive quantization on JPEG, HEVC, and VVC (from Left to right, respectively).

3. Supplementary Experiment Results

3.1. Lo-Hi sensitive compression

We evaluated the compression performance on the testing dataset on JPEG, HEVC, and VVC, wherein four compression conditions for each compression standard are considered. From the rate-distortion curves in Fig. 2, it can be seen that our compression module effectively preserves perceptual quality while significantly saving bitrates compared with the default compression. Meanwhile, BD-rate savings derived from our Lo-Hi compression gradually decreased as the compression computational complexity increased from JPEG to VVC.

3.2. Qualitative results

Comprehensive visual results of compensation frameworks evaluated on JPEG, HEVC, and VVC are showcased in Fig. 5, Fig. 6, and Fig. 7, respectively. Among four compression conditions of each compression standard, our Lo-Hi co-design frameworks consistently offer high perceptual quality in the context of reality, details, and brightness.

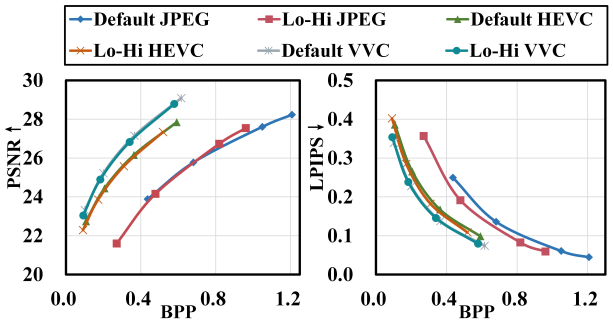


Figure 2. Rate-distortion performance of default compression and Lo-Hi sensitive compression. The corresponding BD rates are reported in Table 2 in Section 5.2.1 of our main manuscript.

(a) HEVC QP45

(b) HEVC QP37

Figure 3. Visual results of object detection evaluated on HEVC at two compression conditions.

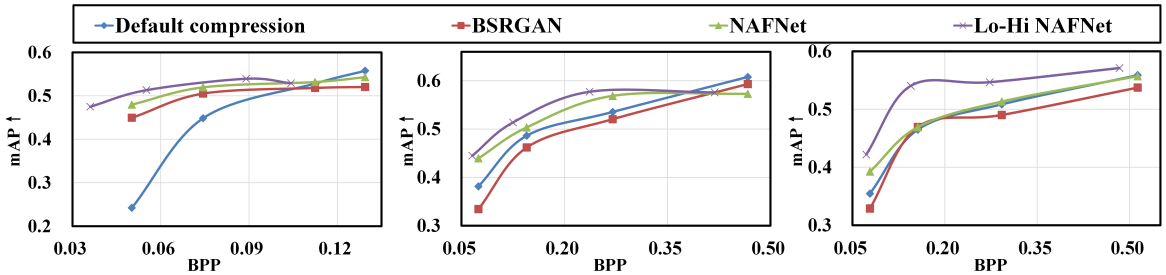


Figure 4. Rate mAP curves under JPEG, HEVC, and VVC compression (from left to right, respectively). Curves are plotted for our Lo-Hi co-design frameworks (Lo-Hi NAFNet), the generative model (BSRGAN with 23 RRDB blocks), the regression model (NAFNet), and the default compression. The corresponding BD rates are reported in Table 4 in Section 5.4 of our main manuscript.

3.3. Machine Vision Results

The experiment of object detection in Section 5.4 of our main manuscript demonstrates that our Lo-Hi co-design framework is robust in machine vision tasks. Rate mAP curves under JPEG, HEVC, and VVC compression are depicted in Fig. 4. It can be seen that our framework achieves remarkable improvement in detection accuracy, especially at low bitrates. Fig. 3 showcases the visual detection results evaluated on HEVC at two compression conditions, wherein default compression, BSRGAN, and NAFNet are compared with our Lo-Hi NAFNet. At the high compression ratio, the "refrigerator" can be correctly detected in our framework while the others give wrong detection results. At the relatively low compression ratio, all images of evaluated frameworks preserve better quality for objective detection. Even though, the "oven" and "refrigerator" can be accurately detected in our framework, while the others fail at detecting "oven" and still provide wrong detection of "refrigerator".

3.4. Complexity of Basis Attention Block

As a plug-in FOI compensation module, BAB is a lightweight network with strong floating-point computation capabilities as shown in Table 1. Compared to increasingly complex regression models, the additional parameters and inference time introduced by BAB are negligible.

Table 1. The complexity of the BAB.

Parameter(M)	Operations (GFLOPs)	Time(s)
0.054	291.19	0.115

References

[1] Eirikur Agustsson and Radu Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2017.

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

[3] 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.

Figure 5. Qualitative comparisons of compensation frameworks on JPEG at four compression conditions. For each sample, QF values of 10, 20, 40, 50 are presented from top to bottom.

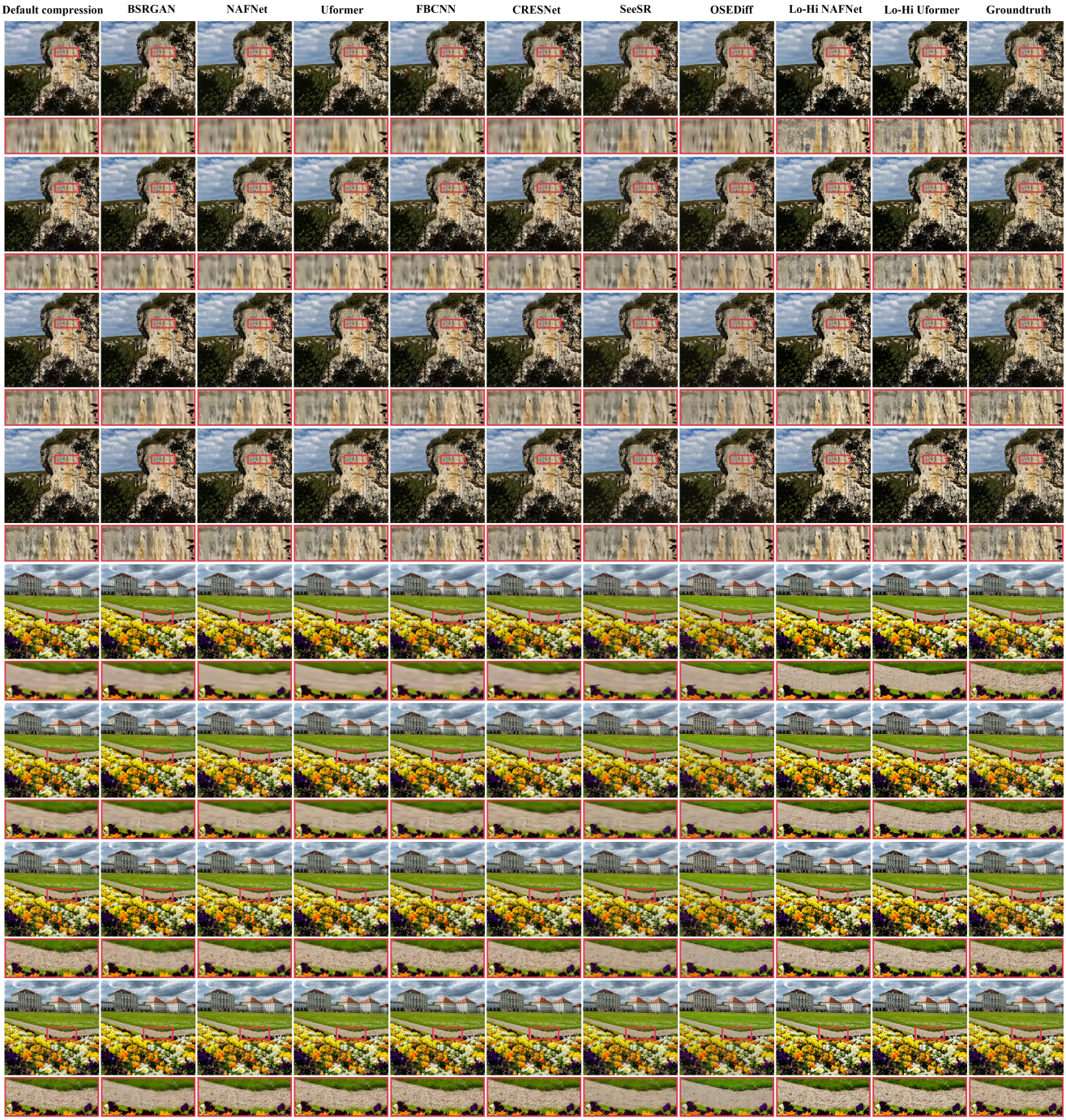


Figure 6. Qualitative comparisons of compensation frameworks on HEVC at four compression conditions. For each sample, QP values of 45, 41, 37, 33 are presented from top to bottom.

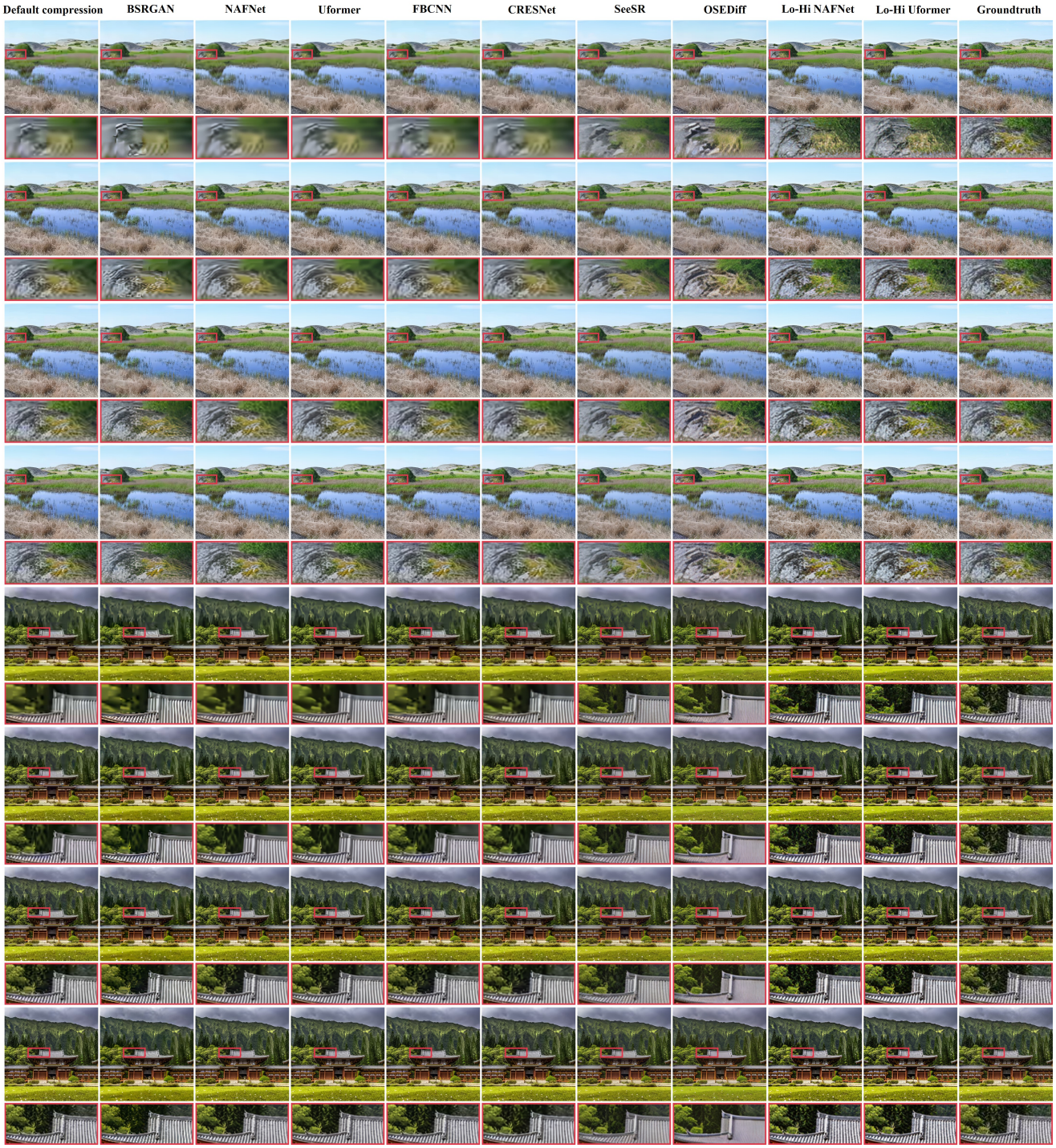


Figure 7. Qualitative comparisons of compensation frameworks on VVC at four compression conditions. For each sample, QP values of 45, 41, 37, 33 are presented from top to bottom.