Supplementary Materials: Consistency Guided Diffusion Model with Neural Syntax for Perceptual Image Compression

Anonymous Authors

1 IMPLEMENTATION DETAILS

1.1 Model Details

The detailed structure of our Consistency Guided Diffusion Model is shown in Table 1 , which is implemented on the basis of the code base IDM [3] with fewer parameters. We are grateful to the developers for their significant contributions.

Table 1: Detailed structure of our Model.

Inner Channel	64
Channel Multiplier	[1,2,2,3]
Depth	2
Dropout	0.2
Attention Resolution	None

1.2 Training Details

For the model training, we adopt the Adam Optimizer [6] with an initial learning rate of 1×10^{-4} and set hyper-parameters to $\beta_1 = 0.9, \beta_2 = 0.999$. Also, any gradients with norm values exceeding 0.5 are clipped to 0.5 or -0.5 to avoid exploding gradients. Besides, we leverage the exponential moving average strategy for more stable training, with decay parameters set to 0.999. The entire training process is conducted on a single NVIDIA GeForce RTX 4090 GPU, using a batch size of 8 during the coarse training stage and a batch size of 1 during the fine training stage.

1.3 Implementation of the compared methods

The code links of all the compared methods are listed in Table 2, we use their officially released pre-trained models for evaluation. And for BPG, we utilize BPG v0.9.8 with the quantizer parameters in the set of [32, 35, 37, 40, 45]. Thanks to the authors for sharing their codes and pre-trained models, which are very helpful for our research work.

1.4 Implementation of the metrics

We implement LPIPS by https://github.com/S-aiueo32/lpips-pytorch/tree/master and other matrics by https://github.com/chaofengc/IQA-PyTorch. When calculating FID [5], the images are cropped to 256×256 patches. We crop all images 2 times from the start position $(0, 0)$ and $(128, 128)$ without overlap.

2 MORE EXPERIMENTAL RESULTS

2.1 Quantitative Comparison

To fully demonstrate the performance of our method, we further show the quantitative performance on the dataset of DIV2K validation [1], which contain 100 high resolution images, by RD-curve in Fig. 1. The performance on DIV2K validation dataset is similar to the performance on Kodak and CLIC professional dataset which are shown in our main paper. Furthermore, we also fully present the BD-rate performance across all evaluation metrics on three datasets, as shown in Table 3, which are partially shown in the main paper.

As can be seen, our approach demonstrates competitiveness in terms of perceptual quality evaluation metrics while surpassing existing state-of-the-art perceptual image compression methods across all distortion metrics.

It is worth noting that, among the perceptual metrics, FID evaluates the distance between the overall distributions of datasets rather than the strict alignment of individual images. Therefore, in compression tasks that prioritize reconstruction fidelity, FID may not be a sufficiently suitable metric for evaluating the performance of image compression methods. Notably, while our method lags slightly behind CDC [11] in the LPIPS metric, CDC's direct utilization of diffusion model significantly compromises the fidelity of its reconstructed images. In contrast, our method maintains competitive perceptual quality while significantly outperforming CDC in terms of distortion, further highlighting the superiority of our consistency-guided approach.

2.2 Qualitative Comparison

More visual results of our CGDM compared with other perceptual image compression methods are shown in Figs. 2, 3, 4 and 5. Fig. 2 and 3 present the visual results on the CLIC professional dataset [10], while Fig. 4 presenting on Kodak dataset [7] and Fig. 5 on DIV2K validation dataset [1]. Evidently, the reconstructed images using our method exhibit richer visual details, and less artifact while utilizing fewer or comparable bits.

Table 2: Code links of the compared methods.

Method	Code Link
BPG [2]	https://bellard.org/bpg/
ELIC [4]	https://github.com/VincentChandelier/ELiC-ReImplemetation
HiFiC [8]	https://github.com/Justin-Tan/high-fidelity-generative-compression
CDC [11]	https://github.com/buggyyang/CDC_compression
ILLM [9]	https://github.com/facebookresearch/NeuralCompression/tree/main/projects/illm

Table 3: Average BD-rate for different methods on three datasets *anchored on our method.*

Datasets	Methods	LPIPS	DISTS	FID	PSNR	VIF	MS-SSIM
Kodak	HiFiC [8]	+88.24	+112.89	+62.70	+29.76	+8.39	+10.70
	ILLM [9]	+30.25	+13.71	+6.71	+12.12	-0.39	+5.13
	CDC [11]	-5.26	+4.50	+10.60	+211.65	+65.05	+90.90
	ELIC [4]	+293.43	+489.23	+547.94	-42.00	-10.45	-12.03
	BPG [2]	+457.11	+702.89	+818.00	-15.22	+3.51	+9.02
	Ours	—	—	—	—	—	—
CLIC	HiFiC [8]	+121.92	+130.91	+88.28	+106.54	+8.46	+43.49
	ILLM [9]	+40.50	+13.70	-11.65	+14.30	+0.23	+6.40
	CDC [11]	-37.44	-11.84	-15.41	+155.01	+9.50	+31.38
	ELIC [4]	+735.35	+1525.43	+4604.38	-41.81	-9.43	-15.45
	BPG [2]	+1023.89	+1810.22	+11782.56	-5.42	+6.34	+8.36
	Ours	—	—	—	—	—	—
DIV2K	HiFiC [8]	+89.46	+117.06	+66.48	+44.58	+10.29	+13.95
	ILLM [9]	+31.57	+11.55	-5.33	+13.15	+0.93	+7.10
	CDC [11]	-17.43	+1.19	+1.53	+198.46	+48.24	+75.49
	ELIC [4]	+313.07	+659.47	+1686.15	-36.65	-10.29	-11.49
	BPG [2]	+484.40	+856.73	+2553.76	+1.40	+6.19	+12.84
	Ours	—	—	—	—	—	—

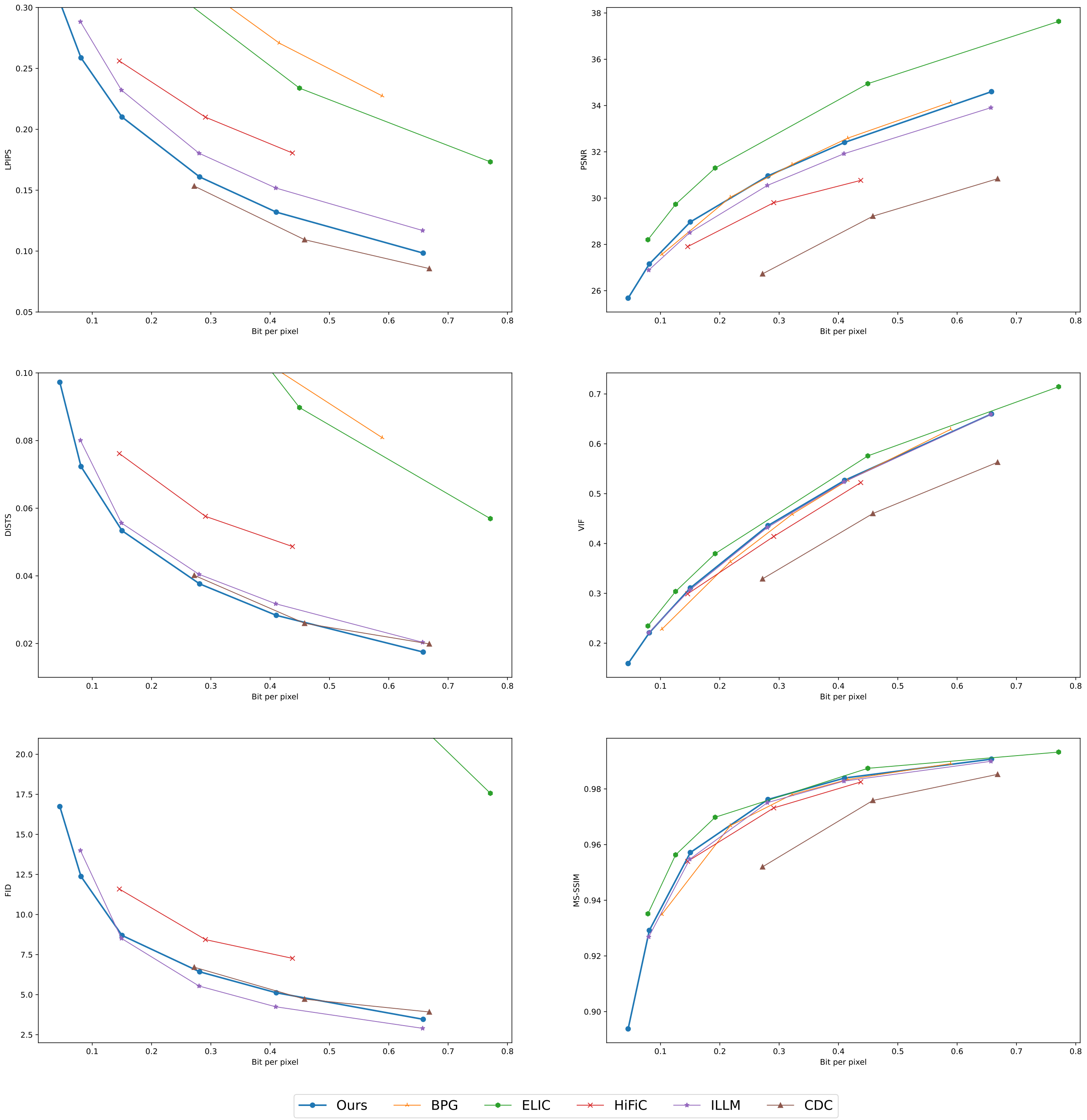
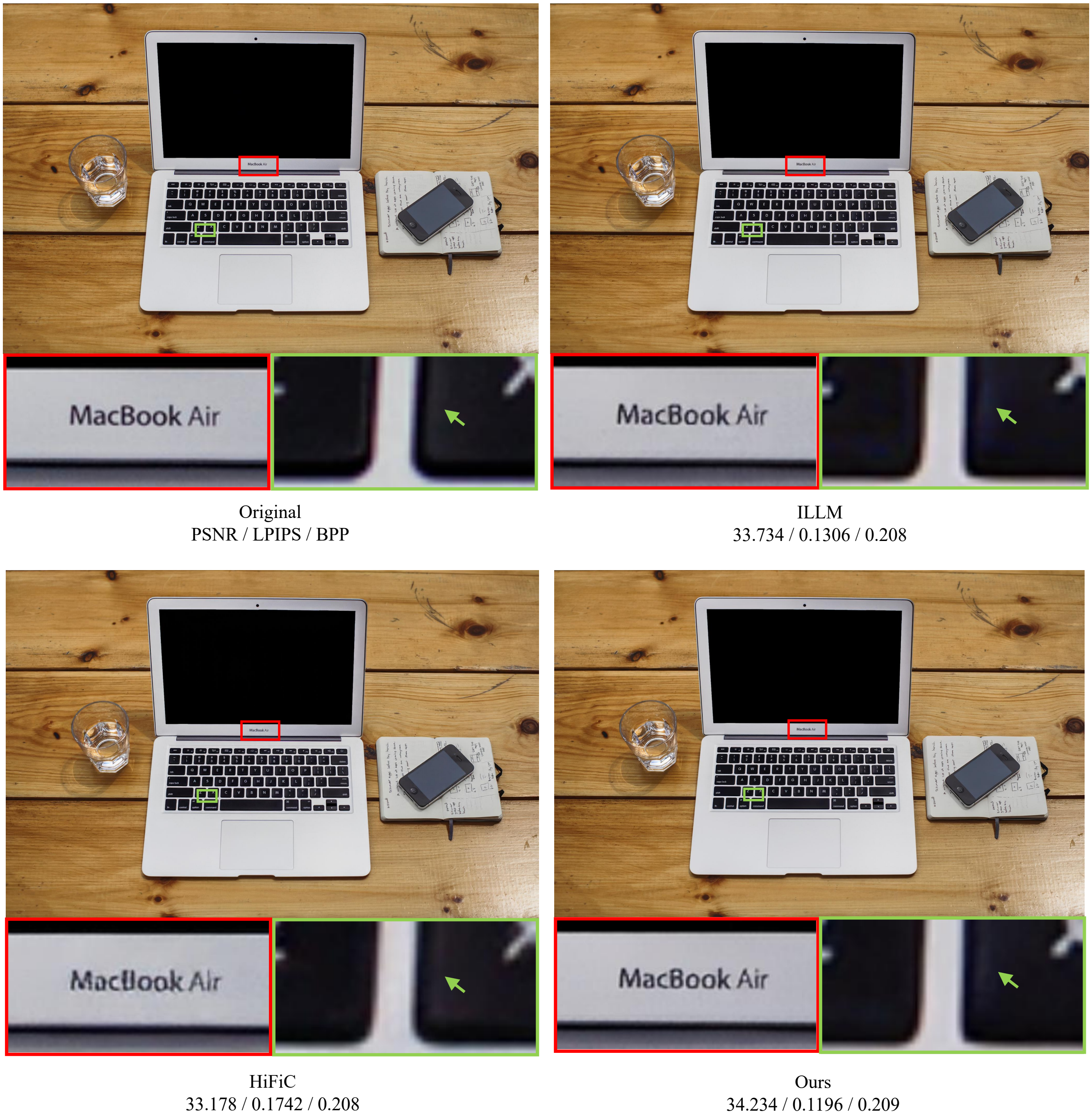


Figure 1: Tradeoffs between bitrate and different metrics for various models tested on DIV2K validation dataset. In the figure above, the left column is related to perceptual quality and the right is related to distortion.[*Zoom in for best view*]



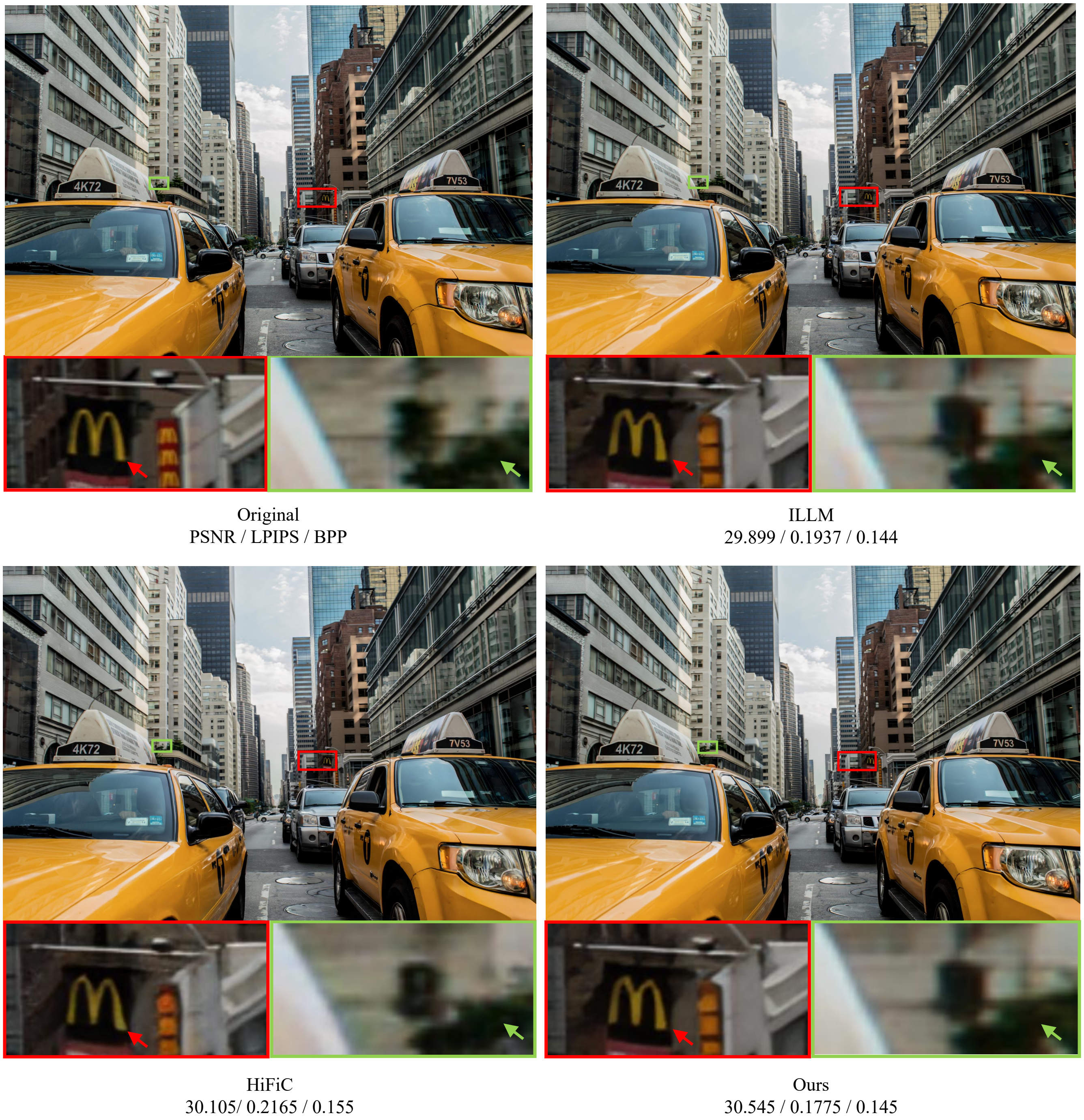
Original
PSNR / LPIPS / BPP

ILLM
33.734 / 0.1306 / 0.208

HiFiC
33.178 / 0.1742 / 0.208

Ours
34.234 / 0.1196 / 0.209

Figure 2: Visual comparisons with state-of-the-art perceptual compression methods on CLIC professional dataset. The red box demonstrates the capability of our method in reconstructing accurate details, while the green box showcases its ability to reconstruct precise colors and textures.[*Zoom in for best view*]



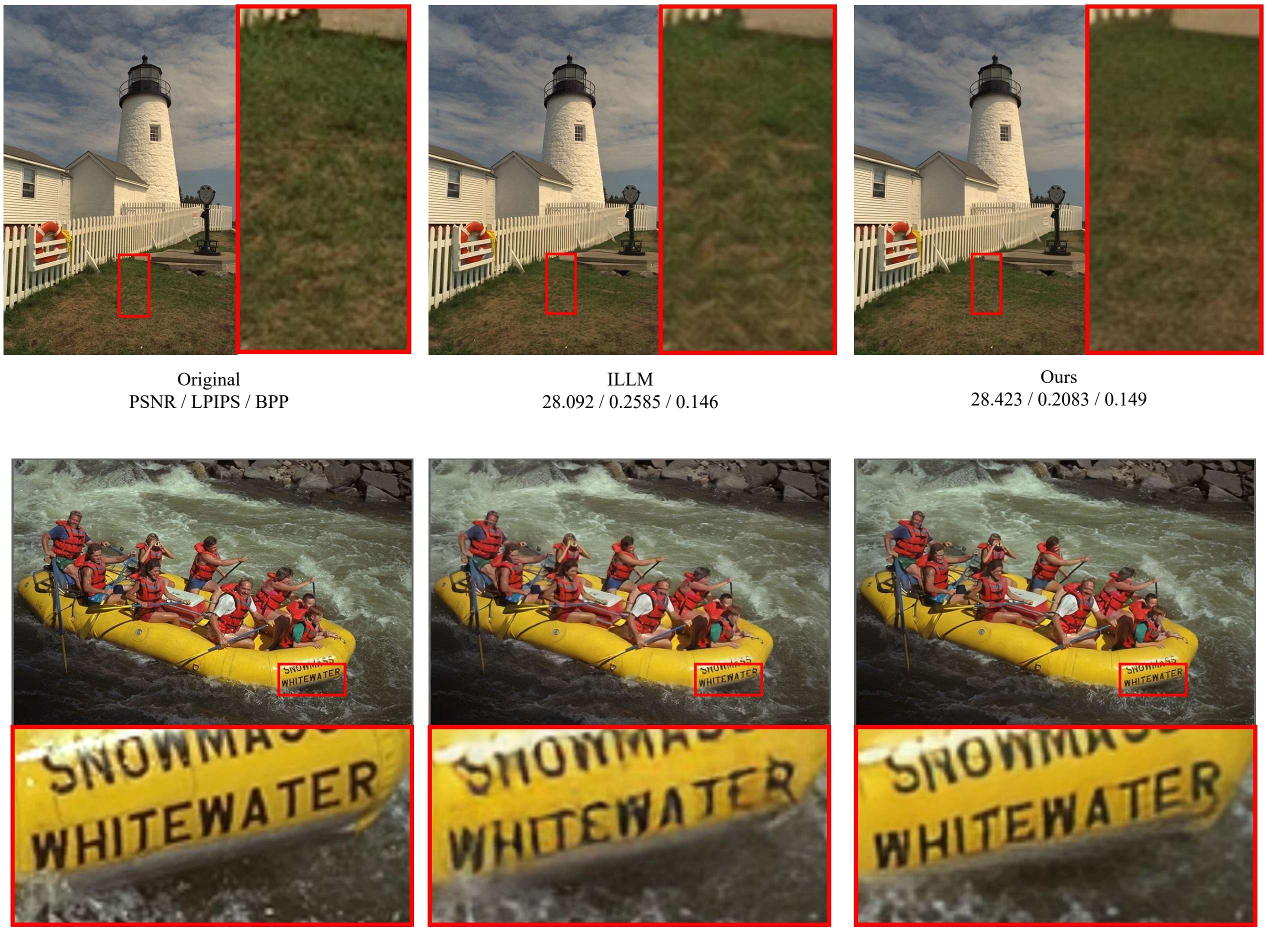
Original
PSNR / LPIPS / BPP

ILLM
29.899 / 0.1937 / 0.144

HiFiC
30.105/ 0.2165 / 0.155

Ours
30.545 / 0.1775 / 0.145

Figure 3: Visual comparisons with state-of-the-art perceptual image compression methods on CLIC professional dataset. [*Zoom in for best view*]



Original
PSNR / LPIPS / BPP

ILLM
28.092 / 0.2585 / 0.146

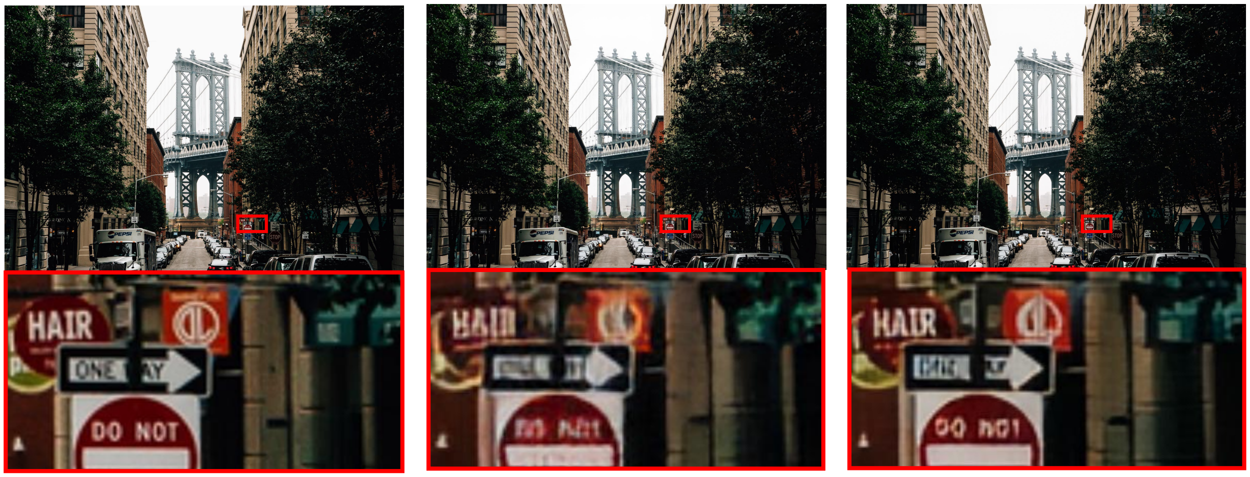
Ours
28.423 / 0.2083 / 0.149

Original
PSNR / LPIPS / BPP

HiFiC
25.530 / 0.3045 / 0.190

Ours
26.324 / 0.2665 / 0.188

Figure 4: Visual comparisons with state-of-the-art perceptual compression methods on Kodak dataset. [*Zoom in for best view*]



Original
PSNR / LPIPS / BPP

HiFiC
26.210 / 0.2431 / 0.408

Ours
27.446 / 0.1852 / 0.372

Figure 5: Visual comparisons with state-of-the-art perceptual image compression methods on DIV2K validation dataset. [*Zoom in for best view*]

REFERENCES

[1] Eirikur Agustsson and Radu Timofte. 2017. NTIRE 2017 challenge on single image super-resolution: dataset and study. In *Proc. IEEE/CVF Int'l Conf. Comput. Vision and Pattern Recognit.*

[2] Fabrice Bellard. 2017. *BPG image format.* http://bellard.org/bpg/

[3] Sicheng Gao, Xuhui Liu, Bohan Zeng, Sheng Xu, Yanjing Li, Xiaoyan Luo, Jianzhuang Liu, Xiantong Zhen, and Baochang Zhang. 2023. Implicit diffusion models for continuous super-resolution. In *Proc. IEEE/CVF Int'l Conf. Comput. Vision and Pattern Recognit.*

[4] Dailan He, Ziming Yang, Weikun Peng, Rui Ma, Hongwei Qin, and Yan Wang. 2022. ELIC: Efficient learned image compression with unevenly grouped space-channel contextual adaptive coding. In *Proc. IEEE/CVF Int'l Conf. Comput. Vision and Pattern Recognit.*

[5] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. 2017. GANs trained by a two time-scale update rule converge to a local nash equilibrium. In *Proc. Annu. Conf. Neural Inf. Process. Systems.*

[6] Diederik Pieter Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).

[7] Eastman Kodak. 2024. *Kodak lossless true color image suite.* https://r0k.us/graphics/kodak/

[8] Fabian Mentzer, George Toderici, Michael Tschannen, and Eirikur Agustsson. 2020. High-fidelity generative image compression. In *Proc. Annu. Conf. Neural Inf. Process. Systems.*

[9] Matthew Muckley, Alaaeldin El-Nouby, Karen Ullrich, Hervé Jégou, and Jakob Verbeek. 2023. Improving statistical fidelity for neural image compression with implicit local likelihood models. In *Proc. Int'l Conf. Mach. Learn.*

[10] George Toderici, Wenzhe Shi, Radu Timofte, Lucas Theis, Johannes Balle, Eirikur Agustsson, Nick Johnston, and Fabian Mentzer. 2020. Workshop and challenge on learned image compression. In *Proc. IEEE/CVF Int'l Conf. Comput. Vision and Pattern Recognit. Workshop.*

[11] Ruihan Yang and Stephan Mandt. 2023. Lossy image compression with conditional diffusion models. In *Proc. Annu. Conf. Neural Inf. Process. Systems.*