One-Step Diffusion for Perceptual Image Compression

Yiwen Jia[1], Hao Wei[1], Yanhui Zhou[2] and Chenyang Ge[1]
[1]Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University, Xi'an, China
[2]School of Information and telecommunication, Xi'an Jiaotong University, Xi'an, China

VCIP 2025
1-4 DECEMBER 2025 | KLAGENFURT, AUSTRIA

Background

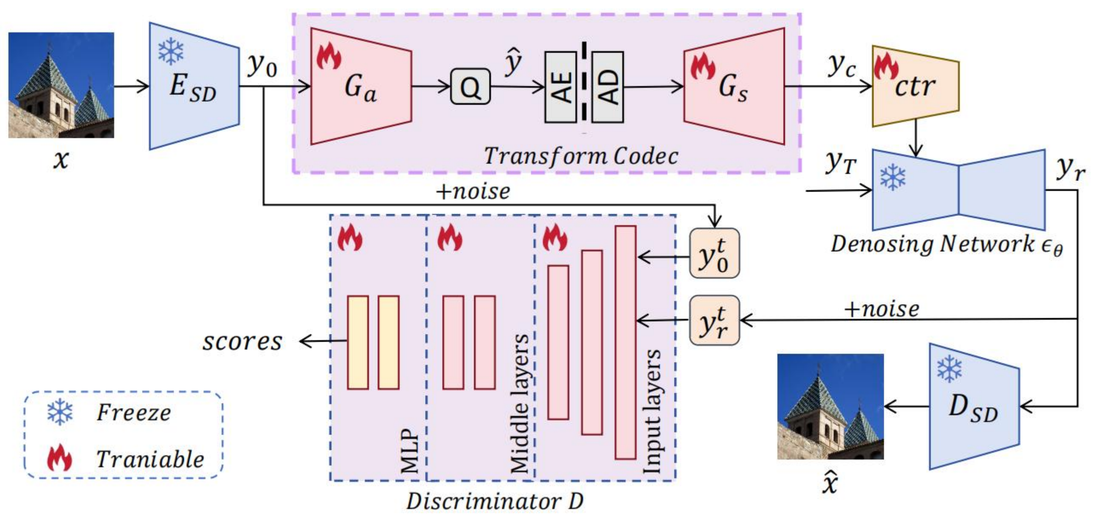
- The demand for digital imagery continues to grow, making efficient **image compression** essential.
  - Traditional codecs (JPEG[1]) rely on hand-crafted heuristics and often cause **visible artifacts**.
  - **Learned image compression** methods (ELIC[2]) improve the rate-distortion trade-off but tend to produce over-smooth reconstructions, especially at low bitrates.

- **Perceptual-driven** generative compression has advanced rapidly with GANs and diffusion models.
- **GAN-based** approaches (HiFiC[3]) suffer from **unrealistic details** and **instability** at low bitrates.
- **Diffusion-based** methods provide more realistic reconstructions, but their multi-step denoising causes **high latency** and **heavy computation**, limiting practical deployment.

Method

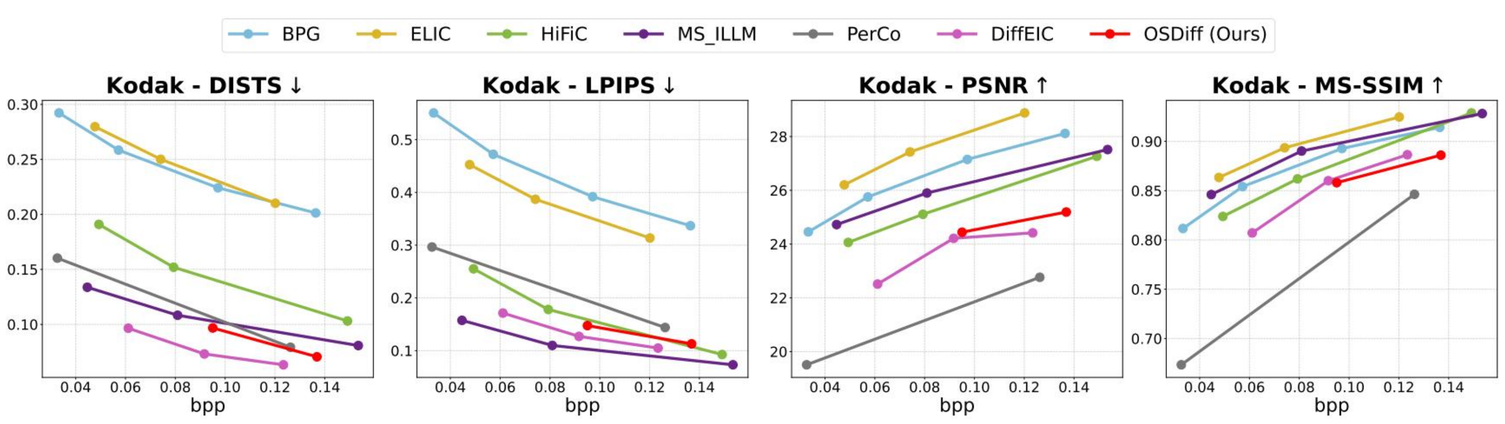
- We introduce **OSDiff**, a **one step** diffusion-based perceptual image compression framework.
- **OSDiff** first encodes the input image, then uses a **control module** to guide the **Denoising Network** to reconstruct a realistic image in one step.
- A **discriminator** compares features from the original and reconstructed images in the middle layers of the **UNet** to enhance perceptual quality without adding inference overhead.

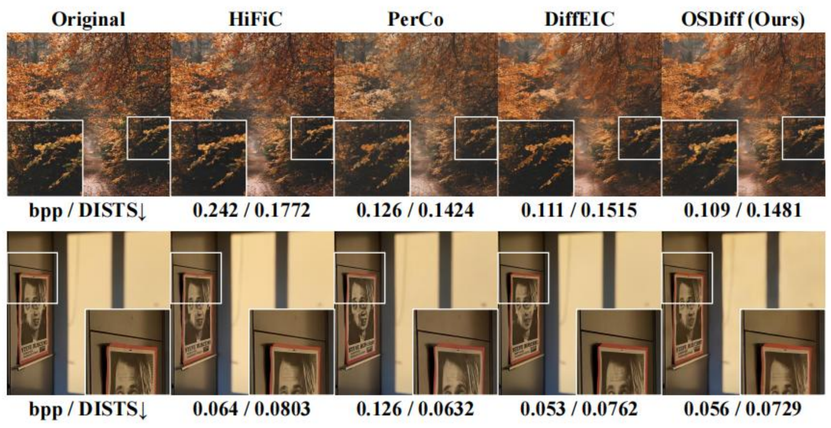
- Our contributions are:
  - We propose a diffusion-based perceptual image compression approach that performs **one-step diffusion**, significantly reducing inference latency and computational cost.
  - We introduce a **discriminator** that operates in a **designated feature space** to further enhance the perceptual quality of reconstructed images.



Results

- Overall, OSDiff achieves the **best performance** on the DISTS metric compared to **other non-diffusion-based methods** and **PerCo**[4].
- Compared to **DiffEIC**[5] with 50-step sampling, OSDiff exhibits a slight drop in perceptual quality due to the one-step sampling strategy. But its **distortion metric** (PSNR) **surpasses** that of **DiffEIC**.





- For an image of size 512×768, OSDiff achieves a decoding time of only **0.060 seconds** on an RTX 4090, which is **approximately 50 times faster** than DiffEIC.

Method	Sampling Steps	Encoding Time (s)	Decoding Time (s)	Device
PerCo	5	0.080	0.665	A100
PerCo	20	0.080	2.551	A100
DiffEIC	50	0.093	2.761	RTX 4090
OSDiff (Ours)	**1**	0.101	**0.060**	RTX 4090

[1] G. K. Wallace, "The jpeg still picture compression standard," Commu_x0002_nications of the ACM, vol. 34, no. 4, pp. 30–44, 1991.
[2] D. He, Z. Yang, W. Peng, R. Ma, H. Qin, and Y. Wang, "Elic:Efficient learned image compression with unevenly grouped space_x0002_channel contextual adaptive coding," in Proceedings of the IEEE/CVF Conference on Computer Vision and
Pattern Recognition, pp. 5718–5727, 2022.
[3] F. Mentzer, G. D. Toderici, M. Tschannen, and E. Agustsson, "High_x0002_fidelity generative image compression," Advances in neural information processing systems, vol. 33, pp. 11913–11924, 2020.
[4] M. Careil, M. J. Muckley, J. Verbeek, and S. Lathuiliere, "Towards image `compression with perfect realism at ultra-low bitrates," in The Twelfth International Conference on Learning Representations, 2023.
[5] Z. Li, Y. Zhou, H. Wei, C. Ge, and J. Jiang, "Towards extreme image compression with latent feature guidance and diffusion prior," IEEE Transactions on Circuits and Systems for Video Technology, 2024.