Steering One-Step Diffusion Model with Fidelity-Rich Decoder for Fast Image Compression

Zheng Chen[1*], Mingde Zhou[1*], Jinpei Guo[2], Jiale Yuan[1], Yifei Ji[1], Yulun Zhang[1†]

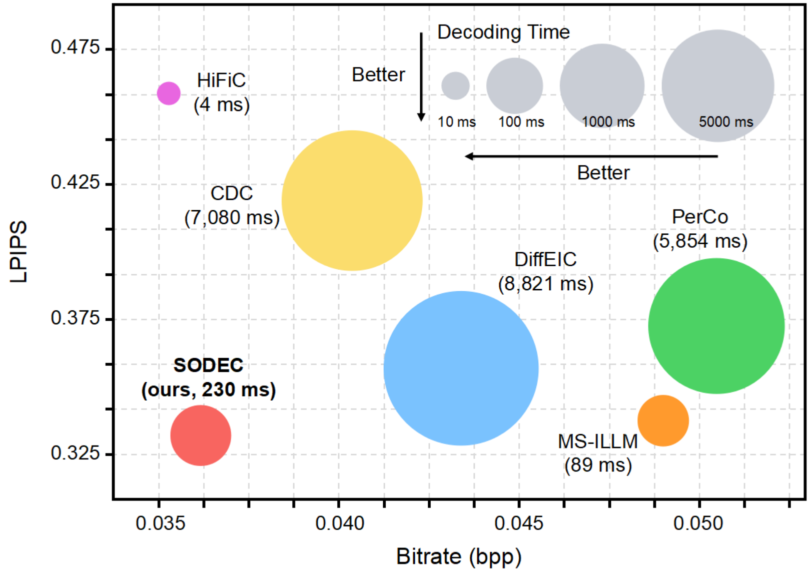
[1]Shanghai Jiao Tong University, [2]Carnegie Mellon University

Introduction

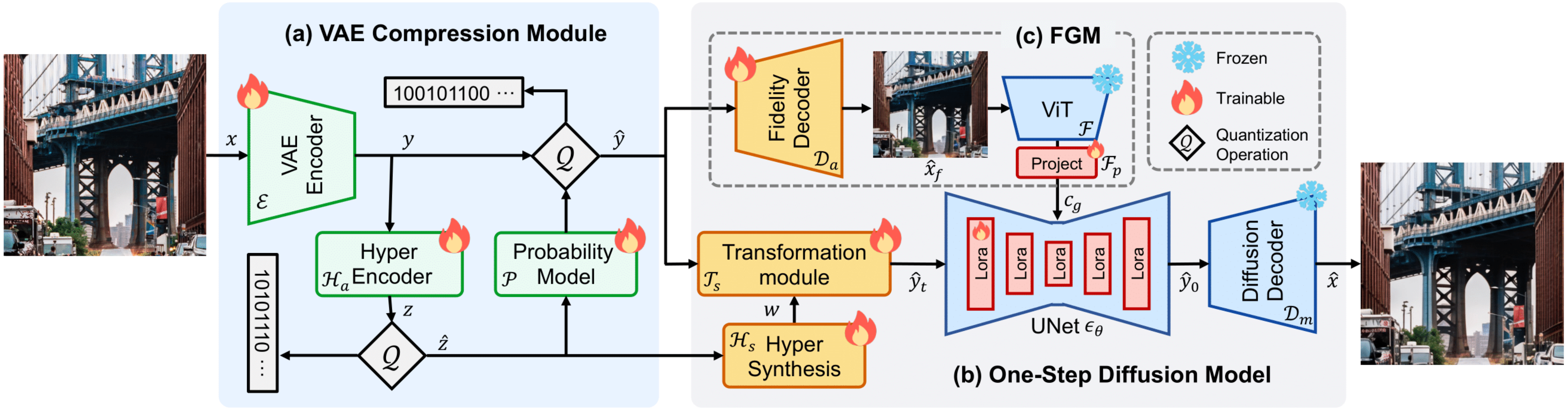
Overview

- Diffusion-based image compression excels under ultra-low bitrates.
- **However:**
  - Existing methods rely on multi-step sampling → high decoding latency.
  - Strong generative priors often cause fidelity deviation from the source image.
- **Goal:** Fast decoding without sacrificing fidelity.

Original HiFiC PerCo DiffEIC SODEC (ours)



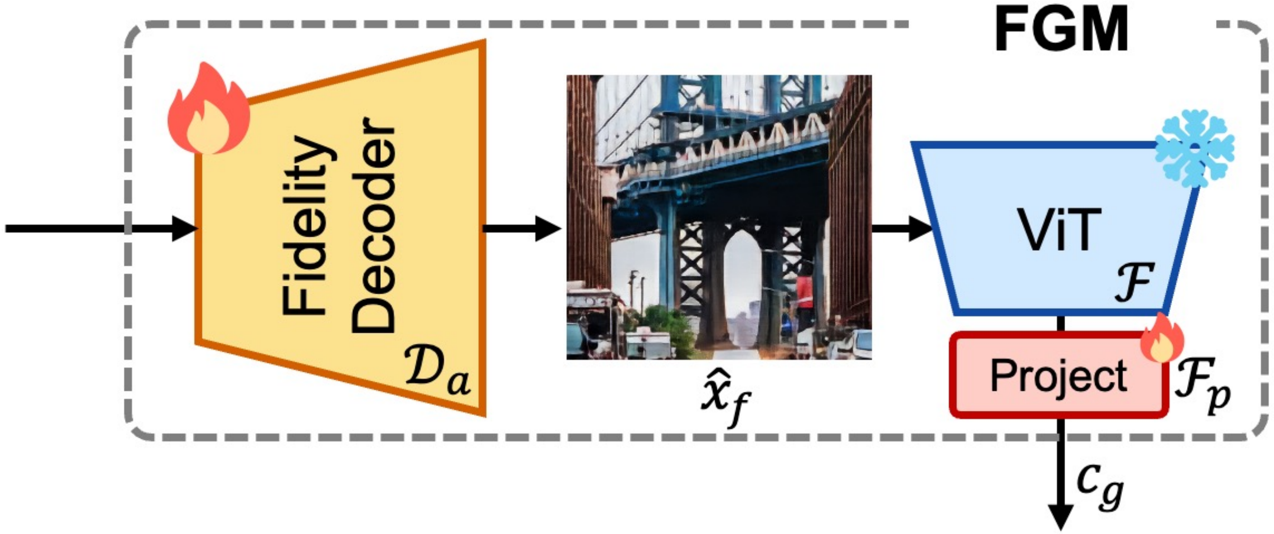
Method

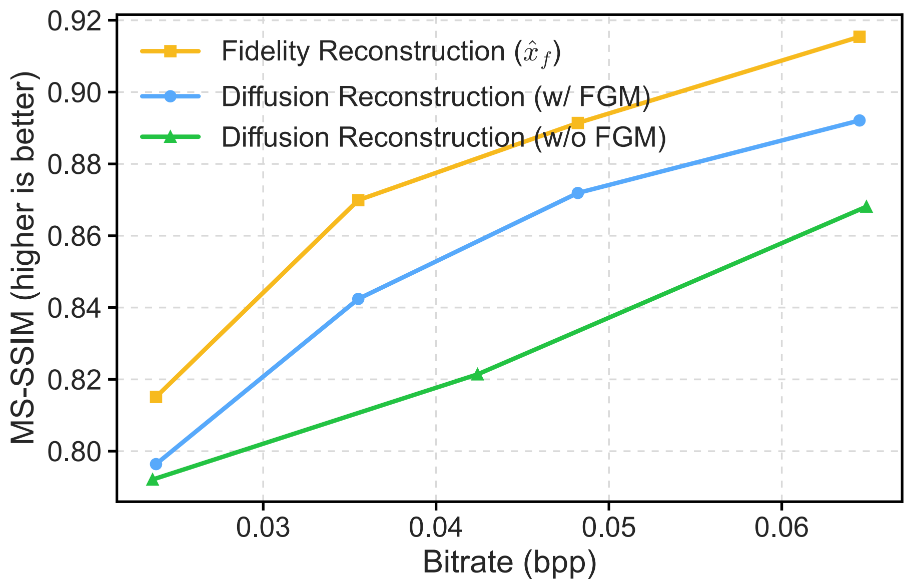


Overview

- VAE-based compression backbone produces informative latent.
- One-step diffusion decoder replaces iterative denoising.

Method





Fidelity Guidance Module

- Generate high-fidelity preliminary reconstruction from VAE decoder.
- Extract visual features as explicit diffusion guidance, and inject guidance via cross-attention.
- Improve content fidelity while preserving perceptual quality.

Experiments

Latency Comparison

- Single-step DM reduces decoding latency by >20×.
- Latency is dominated by one forward pass, no iterative refinement.

Model	Total Time (ms)	Enc. Time (ms) ↓	Dec. Time (ms) ↓	bpp ↓
HiFiC	9.3	5.4	3.9	0.0310
MS-ILLM	9.3	54.5	84.4	0.0395
PerCo	6,242.2	1,540.0	4,702.2	0.0313
DiffEIC	7,827.5	266.4	7,561.1	0.0391
SODEC	232.9	5.0	227.9	0.0314

Guidance Strategy	MS-SSIM ↑	LPIPS ↓	bpp ↓
(i) No Guidance	0.8212	0.3625	0.0424
(ii) Text Prompt Guidance	0.8185	0.3631	0.0412
(iii) Hyperprior Guidance	0.8258	0.3527	0.0385
(iv) Aux. Fidelity Guidance (ours)	0.8481	0.3351	0.0368

FGM

- FGM significantly improves fidelity.
- Explicit visual guidance is more effective.

Experiments

Alignment Loss Config.	MS-SSIM ↑	LPIPS ↓	bpp ↓
(i) No Alignment Loss	0.7490	0.4210	0.0203
(ii) MSE + LPIPS	0.7481	0.3961	0.0199
(iii) Merged into Main Loss	0.7984	0.4023	0.0232
(iv) MSE only (ours)	0.7948	0.3827	0.0227

Alignment Loss

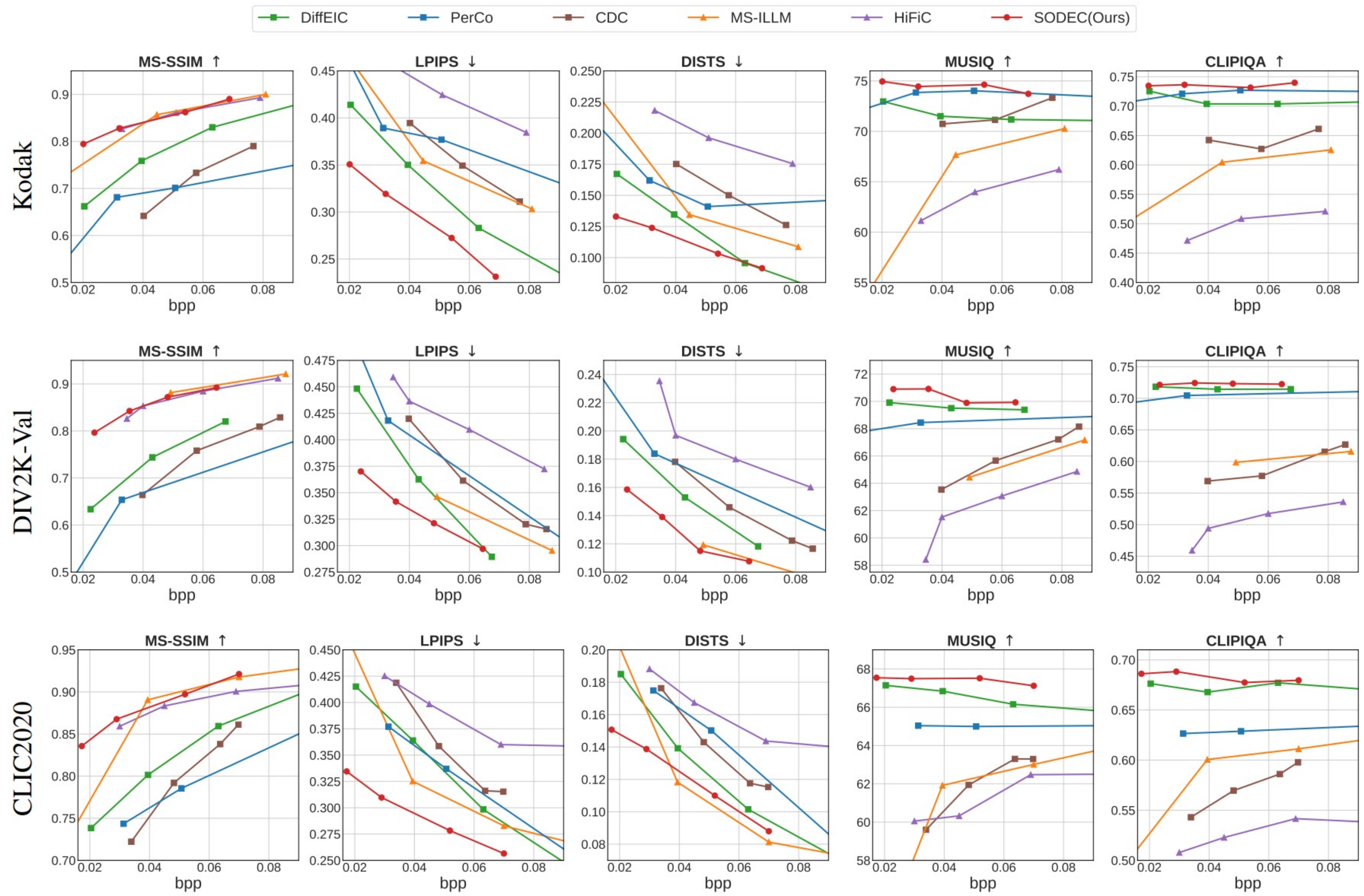
- Alignment loss is necessary to maintain high-fidelity.
- MSE-only achieves the best fidelity–perception balance.

Training Strategy

- Rate annealing outperforms direct joint training at matched bitrates.
- High-to-low bitrate training enables better selection.

Training Strategy	MS-SSIM ↑	LPIPS ↓	bpp ↓
(i) Frozen VAE Module	0.8512	0.3761	0.0695
(ii) Joint Training (Matched bpp)	0.8621	0.3750	0.0678
(iii) Low-to-High bpp Curriculum	0.8643	0.3451	0.0593
(iv) Rate Annealing (ours)	0.8951	0.3113	0.0604

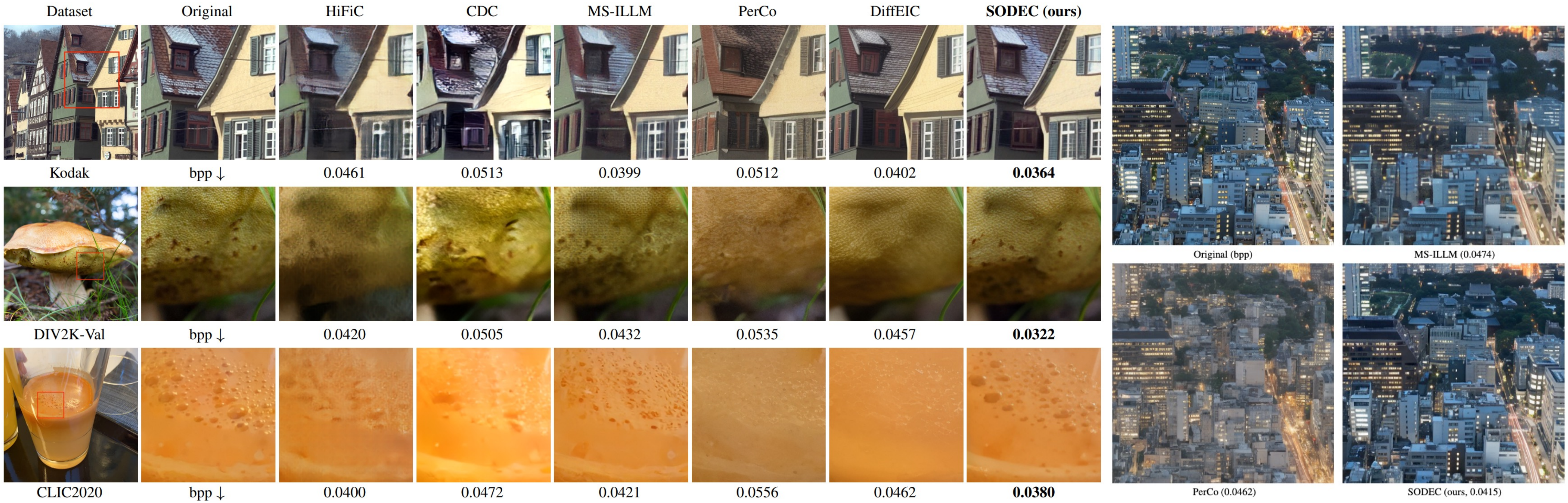
Experiments



Quantitative

- SODEC achieves state-of-the-art performance across different datasets,
- Outperforms multi-step diffusion methods in perceptual metrics
- Over **20×** faster than diffusion-based competitors

Method



Qualitative

- SODEC reconstructs images with more accurate structures and fewer artifacts.

Conclusion

Contribution

- Propose SODEC, a steered one-step diffusion model for fast image compression.
- Introduce a fidelity-rich decoder to guide diffusion toward faithful reconstruction.
- Design a rate annealing training strategy for optimization at ultra-low bitrates.
- Achieve state-of-the-art performance with >20× decoding speedup.

Project

Home Page

Thanks!