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

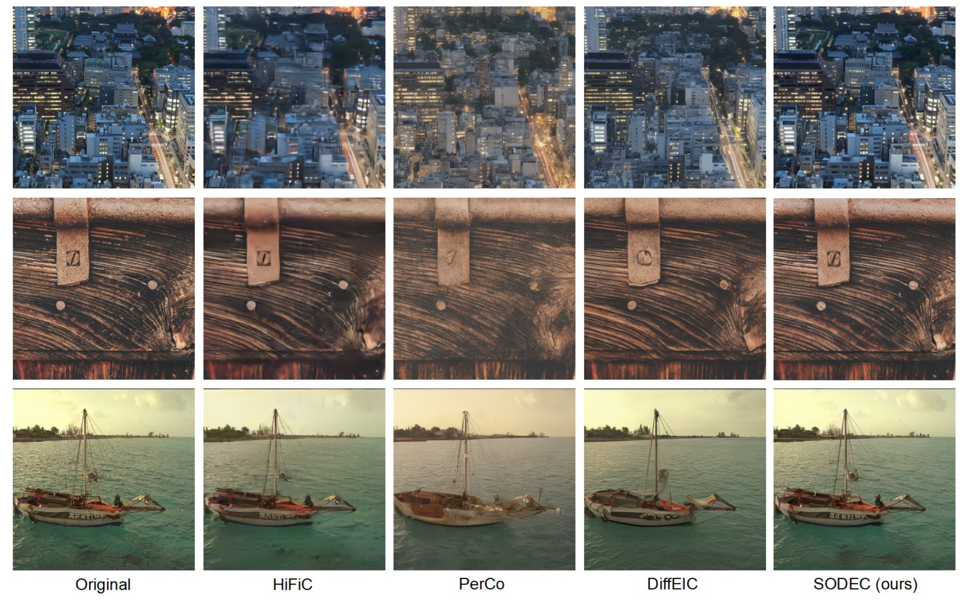
Zheng Chen[1*], Mingde Zhou[1*], Jinpei Guo[2], Jiale Yuan[1], Yifei Ji[1], Yulun Zhang[1†]

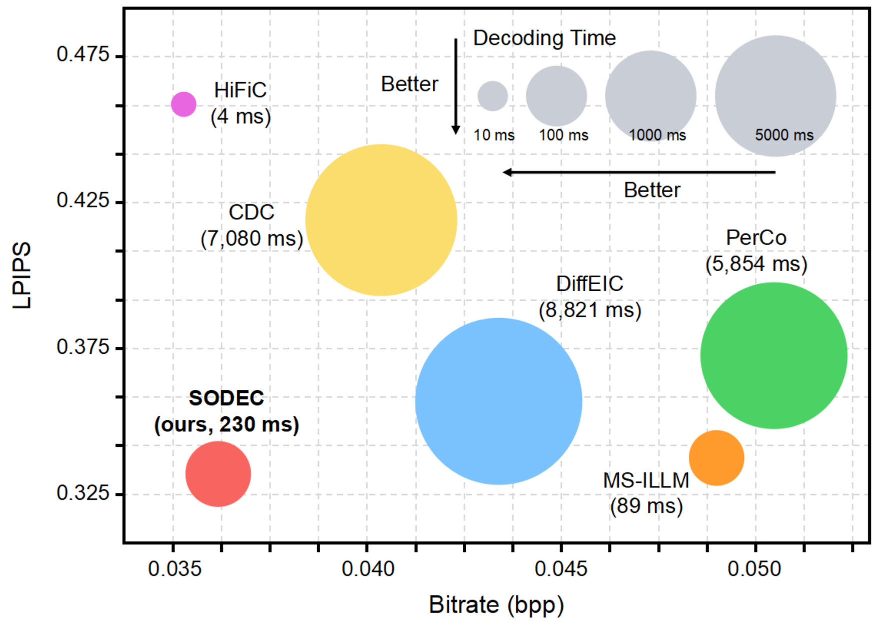
[1]Shanghai Jiao Tong University, [2]Carnegie Mellon University

Project   Home Page

Introduction

- 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.



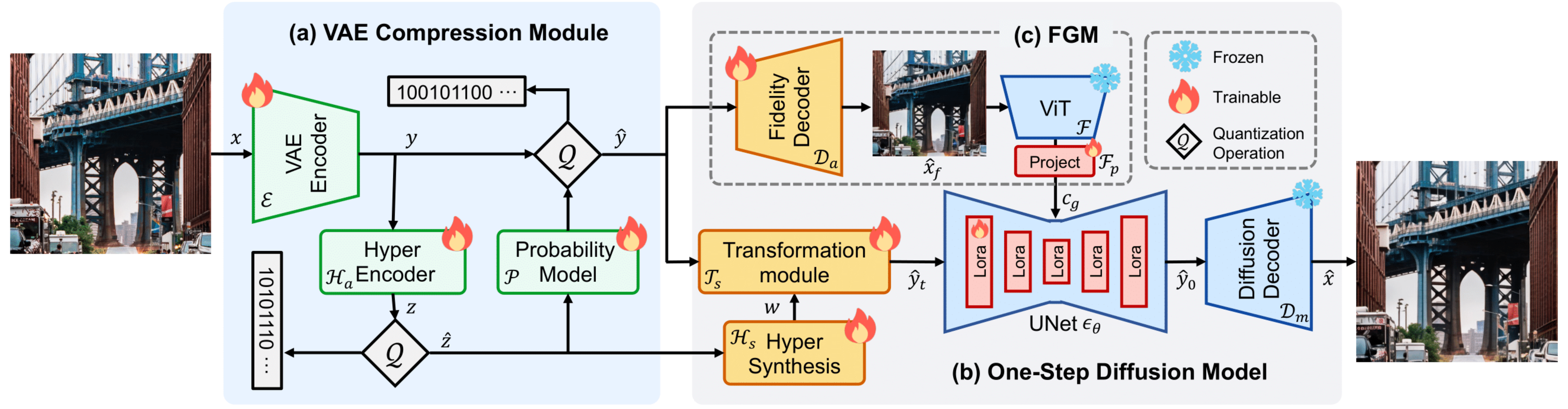


Contribution

- 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.

Method

❖ Overall



❖ SODEC Framework

The framework consists of three key components:

- VAE-based compression module producing latent.
- one-step diffusion decoder for fast reconstruction.
- fidelity guidance module that injects high-fidelity cues.

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

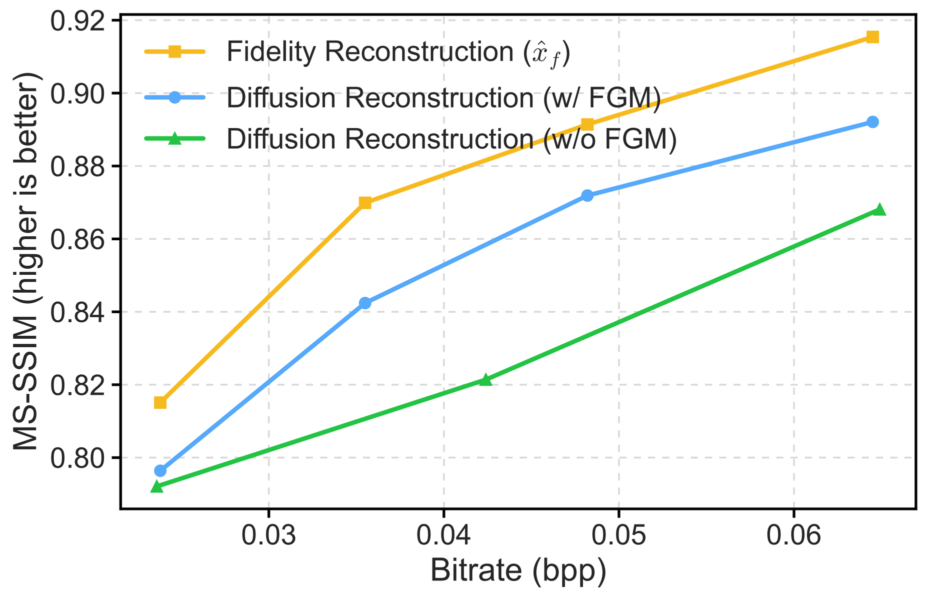
Table 1: Inference efficiency comparison on the DIV2K-Val dataset. Total, encoding, and decoding times are measured on one A6000 GPU with the 512×512 image.

❖ One-Step Diffusion Model

- SODEC performs single-step denoising at a fixed timestep.
- The informative latent representation makes iterative refinement unnecessary.

❖ Fidelity Guidance Module

- Generate high-fidelity preliminary reconstruction from VAE decoder.
- Extract visual features as diffusion guidance.
- Inject guidance via cross-attention.



❖ Rate Annealing Training Strategy

- **Stage 1:** High-Bitrate VAE Pre-training.
- **Stage 2:** Diffusion Path Warm-up.
- **Stage 3:** Joint Training with Rate Annealing.

Experiments

❖ Ablation Study

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

Table 2: Ablation on the fidelity guidance module.

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

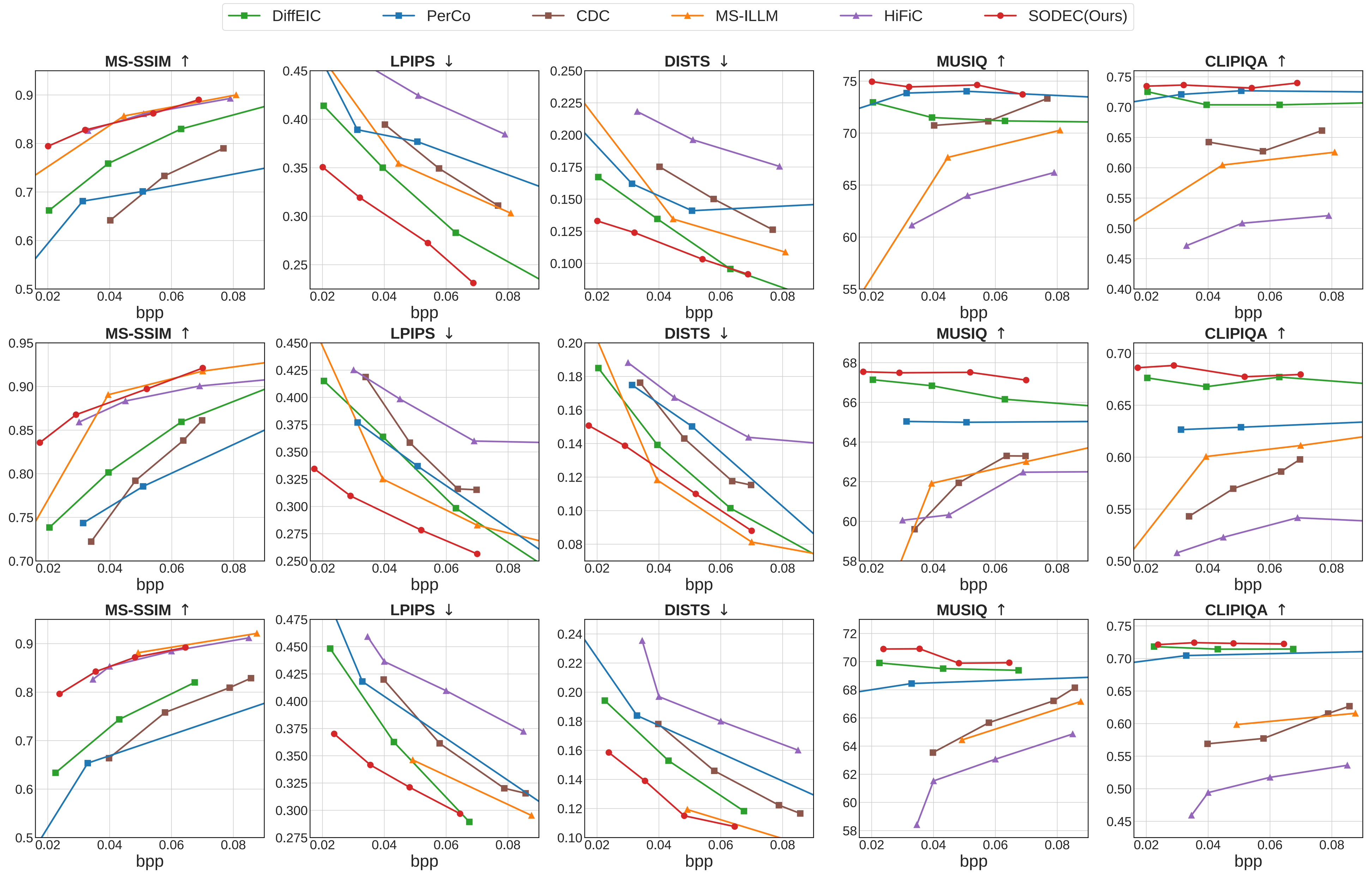
Table 3: Ablation on the setting of alignment loss ($\mathcal{L}_{align}$).

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

Table 4: Ablation study on different training strategies.

Experiments

❖ Quantitative Results



❖ Qualitative Results