COSMIC: Compress Satellite Images Efficiently via Diffusion Compensation

Ziyuan Zhang[1], Han Qiu[1,#], Maosen Zhang[1], Jun Liu[1], Bin Chen[2], Tianwei Zhang[3], Hewu Li[1]
[1]Tsinghua University, China
[2]Harbin Institute of Technology, Shenzhen, China
[3]Nanyang Technological University, Singapore
{ziyuan-z23,zhangms24}@mails.tsinghua.edu.cn, {qiuhan,juneliu}@tsinghua.edu.cn
chenbin2021@hit.edu.cn, tianwei.zhang@ntu.edu.sg, lihewu@cernet.edu.cn

Outline

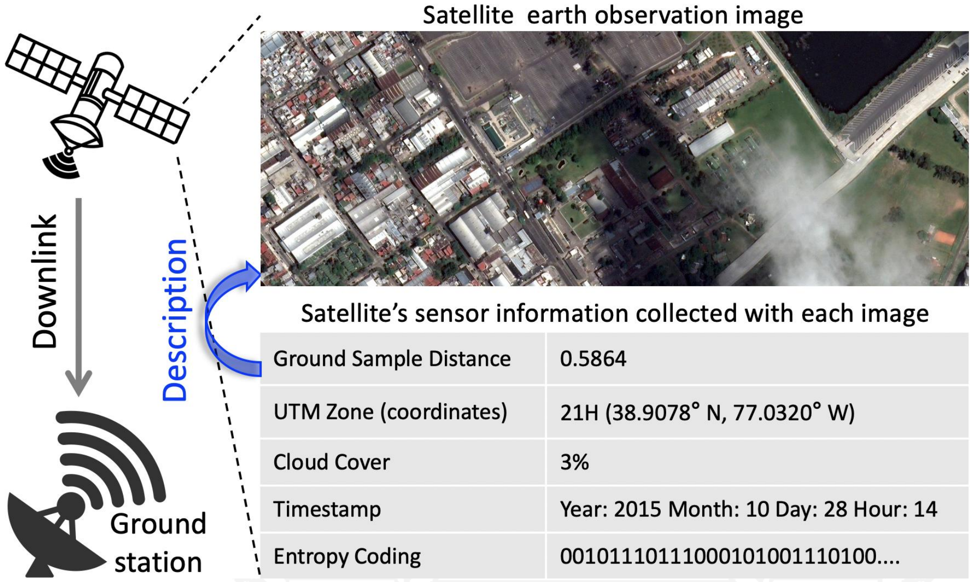
- Background and motivation
- The proposed method
- Experiment
- Conclusion

Background and motivation

- Images collected by satellites grow rapidly
  - E.g. Sentinel-3 missions can collect maximum 20 TB raw data every day
- The data transmission capability between satellites and ground stations has clear upper bounds
- **Downlink Bottleneck -> effectively compress satellite images before transmission**
- A novel promising trend: deploying embedded GPUs on satellites
  - E.g. VPU on Phi-Sat-1 and Jetson Xavier NX on Forest-1
  - Limited computing capacity: 21TOPS (NX) v.s. 285TOPS (3090)
- **Computational bottleneck -> hard to depoly existing learned compression solutions**

Background and motivation

- We fill the above gaps from two aspects
- Computational: a lightweight encoder
- Image quality: a sophisticated decoder with a diffusion based compensation module



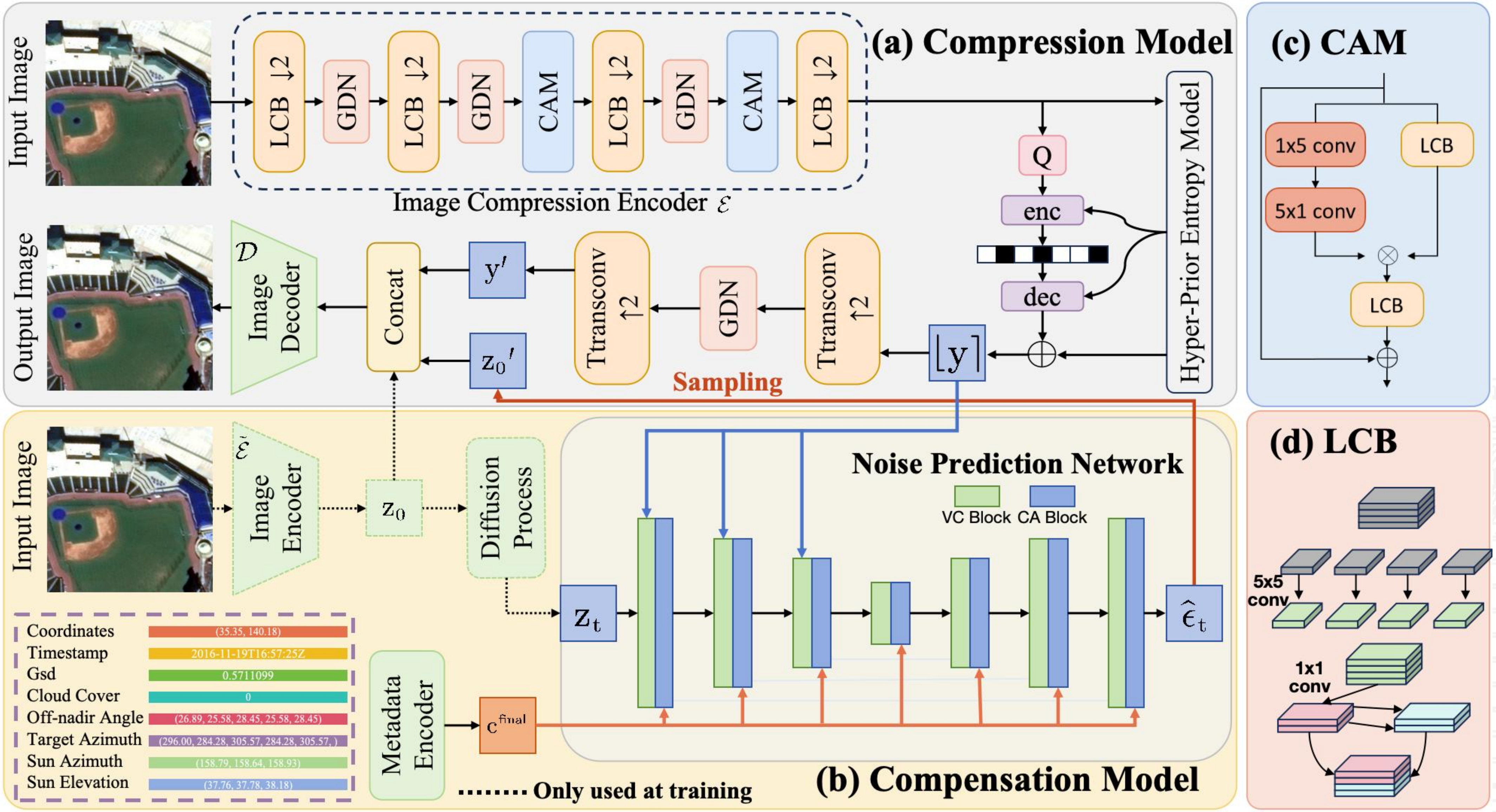
Key insight !!

Satellites' earth observation photos enjoy a multi-modal nature in which rich real-time sensor information is the description of the corresponding photo.

Outline

- Background and motivation
- **The proposed method**
- Experiment
- Conclusion

The proposed method



The proposed method

- Lightweight compression encoder
  - LCB block and CAM block
  - CAM block
    - Two one-dimensional convolutions in series
    - Capture the global attention map
  - LCB block
    - Leverage redundancy between feature map

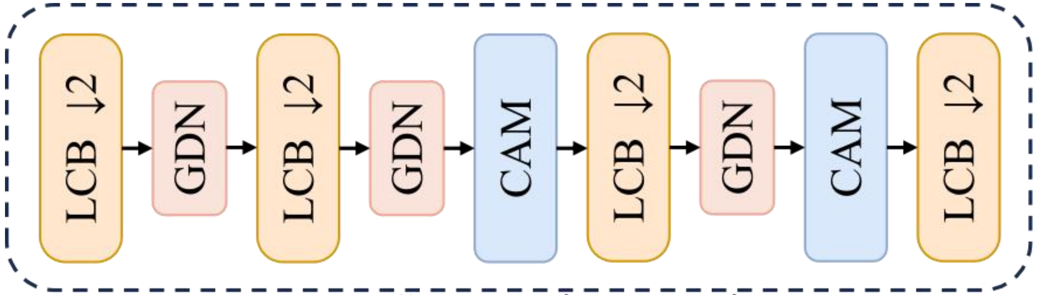
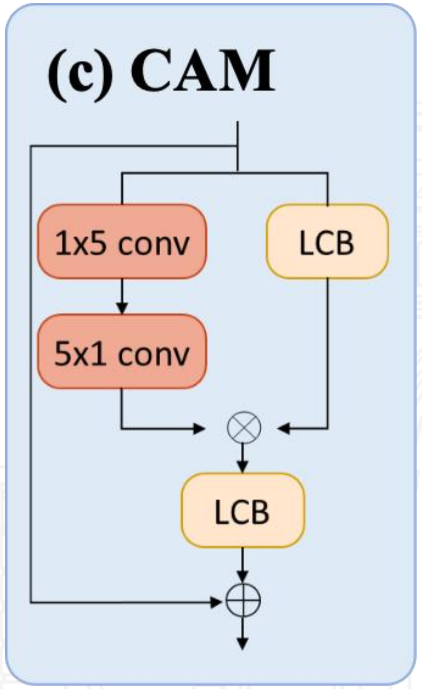
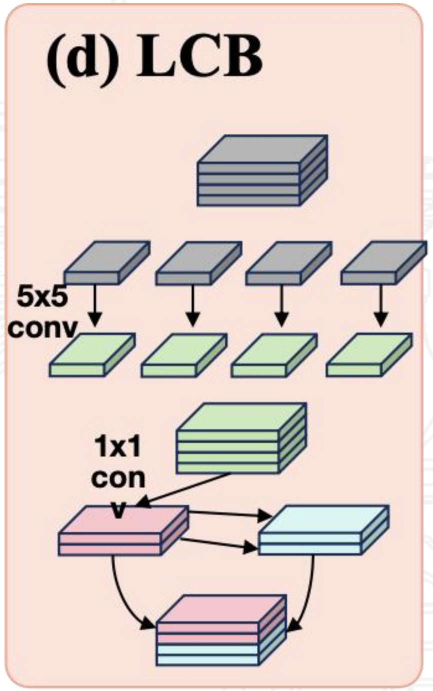


Image Compression Encoder $\mathcal{E}$

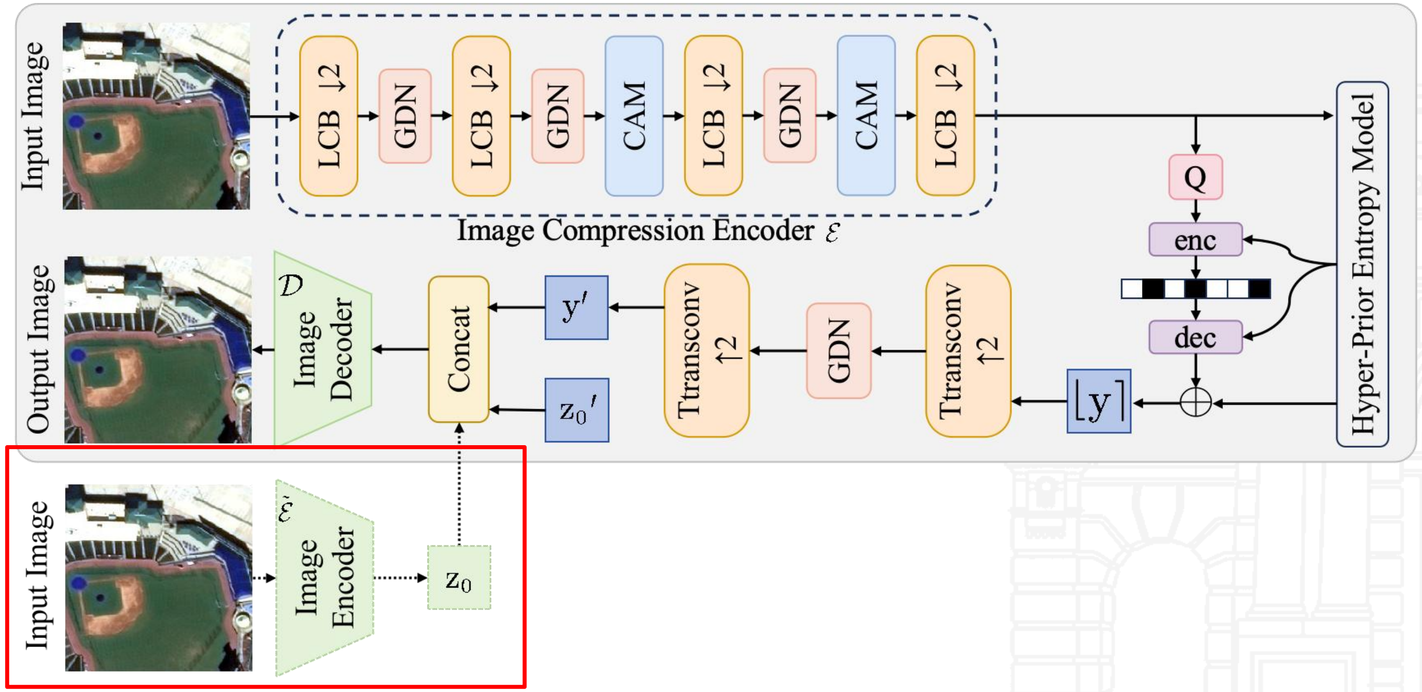




The proposed method

- Compensation-guided image compression
  - image encoder $\tilde{\mathcal{E}}$ -> extract compensation information
  - First stage training: train $\mathcal{E}$ $\tilde{\mathcal{E}}$ and $\mathcal{D}$ together

$$\mathcal{L}_{\mathrm{IC}} = \mathrm{R} + \lambda \mathrm{D} = \mathbb{E}\left[-\log_2 \mathrm{p}\left(\lfloor \mathbf{y} \rceil | \zeta\right) - \log_2 \mathrm{p}\left(\zeta\right)\right] + \lambda \mathbb{E}\left[\mathrm{d}\left(\mathbf{x}, \hat{\mathbf{x}}\right)\right]$$



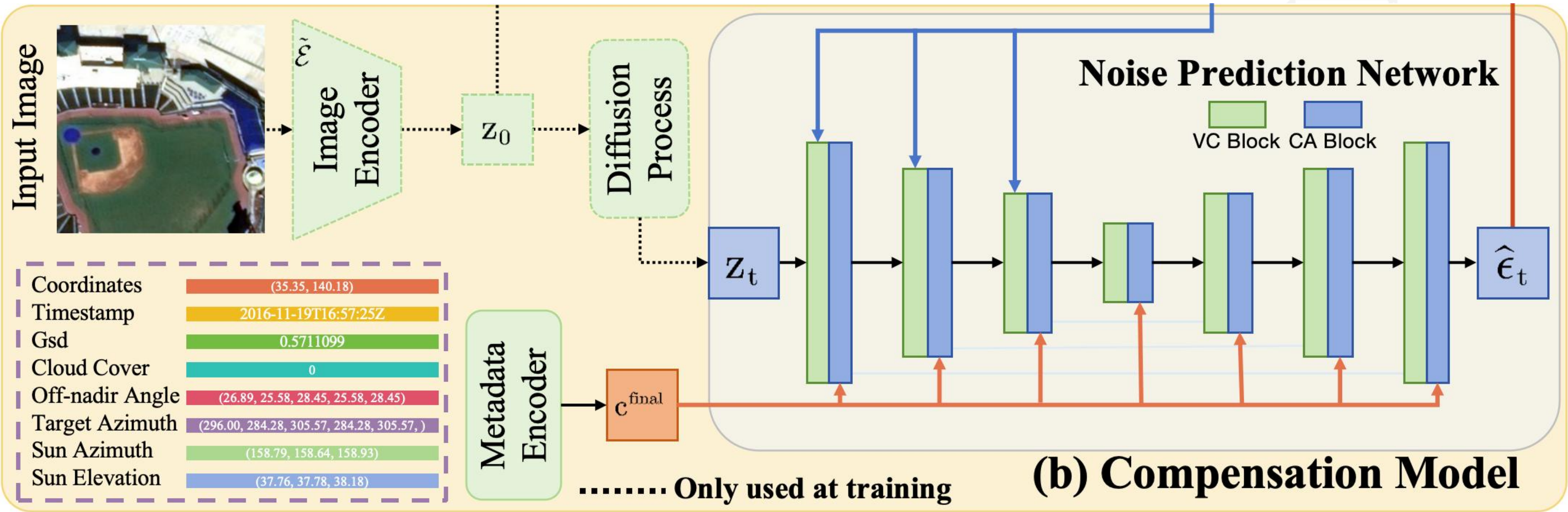
The proposed method

- Conditional diffusion model for loss compensation
  - Metadata projection

$$c^{\text{final}} = \text{MLP}\left(\text{concat}\left(\left[E_{\sin}(m_1), \ldots, E_{\sin}(m_M)\right]\right)\right)$$

  - Second stage training: finetune diffusion model

$$\mathcal{L}_{\text{ldm}} = \mathbb{E}_{t, z_0, \epsilon \sim \mathcal{N}(0,1)}\left[\left\| \epsilon_t - \epsilon_\theta\left(z_t, t, \lfloor \mathbf{y} \rceil, c^{\text{final}}\right) \right\|^2\right]$$

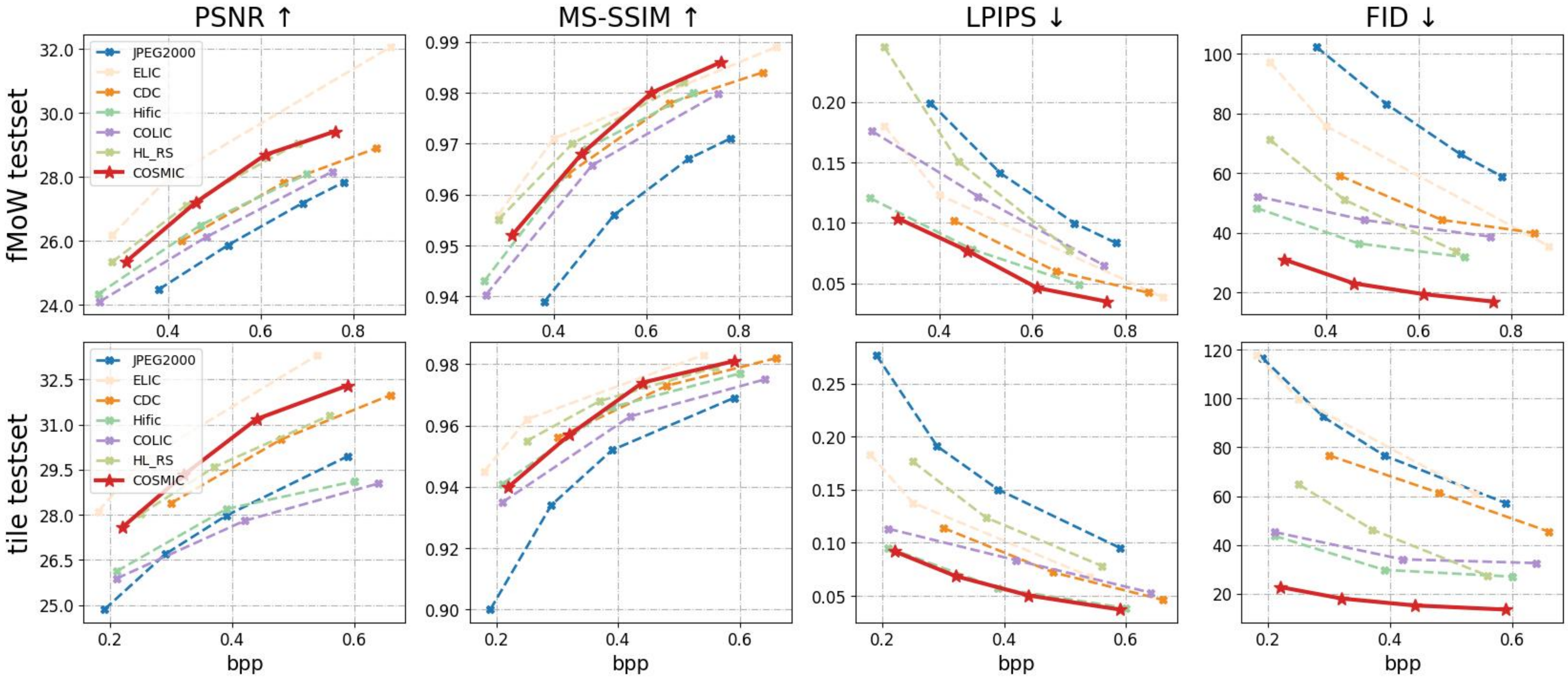


Outline

- Background and motivation
- The proposed method
- Experiment
- Conclusion

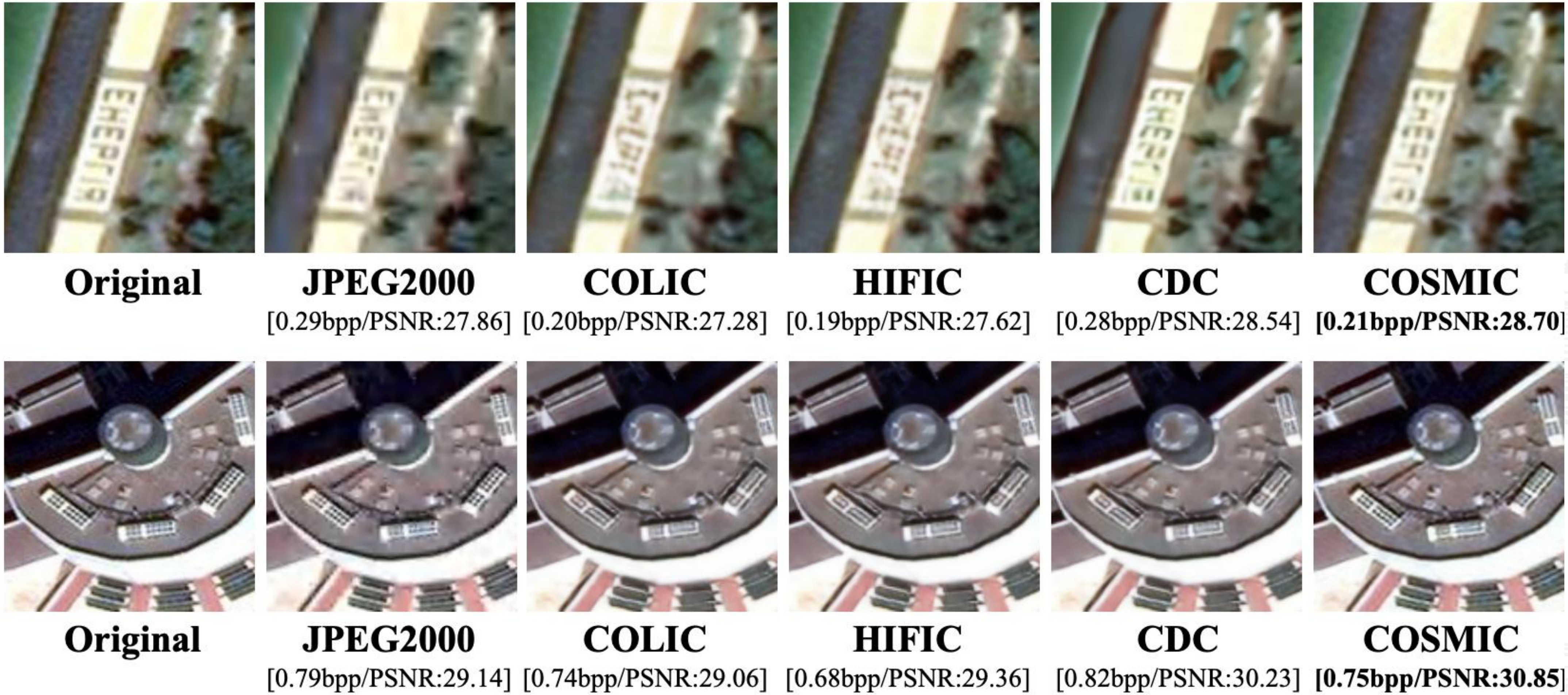
Experiments

- Quantitive results



Experiments

- Visual results



Experiment

- Visual results -- tiling scenarios

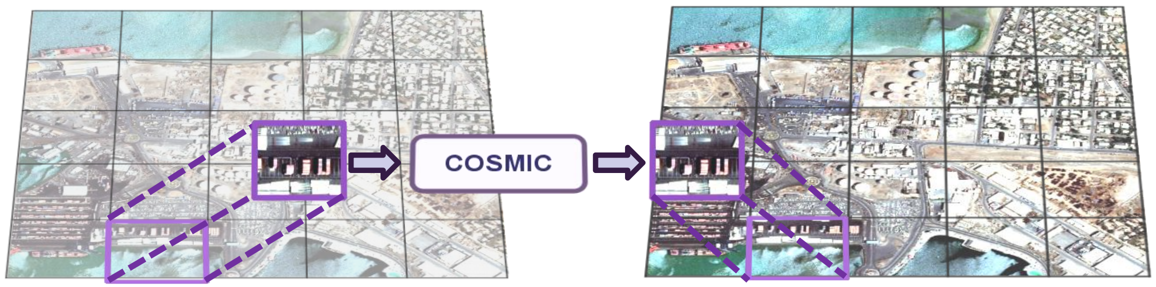
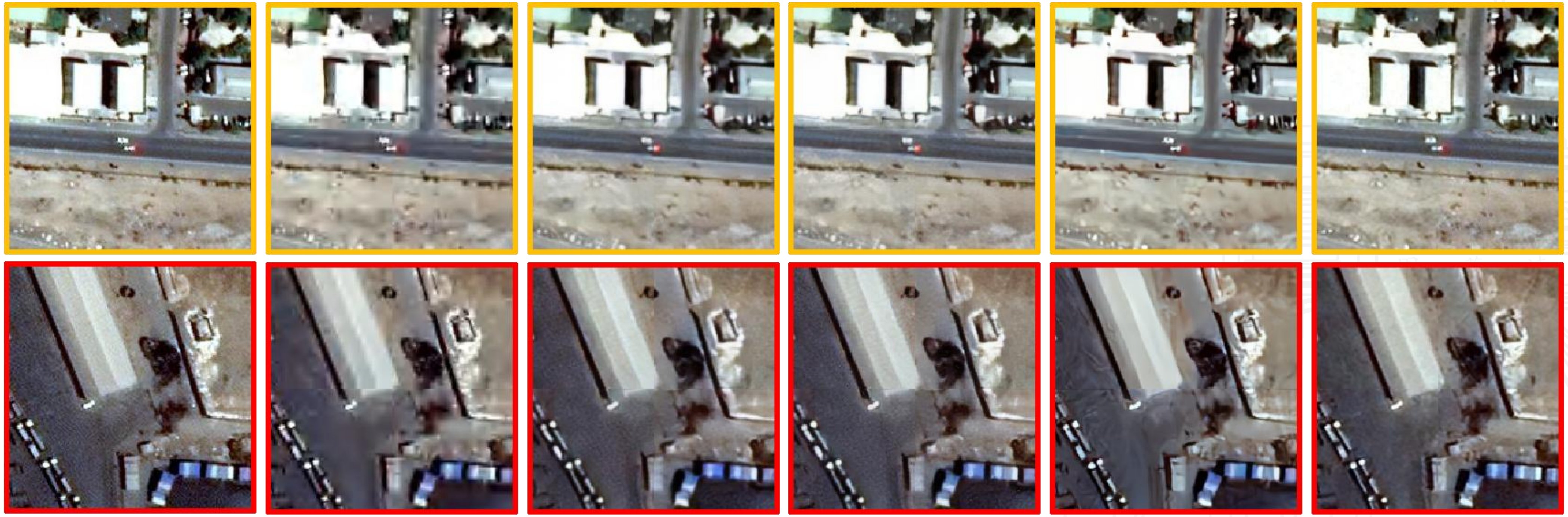


Illustration for tile testset



Original	JPEG2000	COLIC	HIFIC	CDC	COSMIC
	[0.38bpp/PSNR:24.41]	[0.24bpp/PSNR:23.77]	[0.24bpp/PSNR:24.04]	[0.41bpp/PSNR:24.82]	[0.30bpp/PSNR:24.75]

Experiment

- Compression influence on downstream tasks

Classes	Original	JPEG2K [46]	Elic [28]	COLIC [36]	HIFIC [39]	CDC [53]	HL_RS [50]	COSMIC
10	98.95%	-4.21%	-5.27%	-1.06%	-1.06%	-2.11%	-2.11%	-1.06%
15	97.92%	-4.17%	-3.84%	-0.70%	-0.70%	-1.40%	-1.39%	-0.70%
20	98.42%	-3.16%	-3.68%	-0.53%	-0.53%	-1.06%	-2.1%	-1.06%

- Encoder efficiency

Method	FLOPs (G)	PSNR↑	MS-SSIM↑	LPIPS↓	FID↓	bpp↓
CDC [53]	13.1	31.98	0.982	0.0462	45.49	0.66
COLIC [36]	26.4	29.04	0.975	0.0530	32.56	0.64
Hific [39]	26.4	29.11	0.977	0.0384	27.01	0.60
Elic [28]	21.78	**33.31**	**0.983**	0.0683	60.80	0.54
HL_RS [50]	11.87	31.30	0.979	0.0782	27.38	0.56
COSMIC	**4.9**	32.11	0.980	**0.0359**	**13.50**	0.59

Achieving SOTA performance with 2.6~5x lower FLOPs

Outline

- Background and motivation
- The proposed method
- Experiment
- Conclusion

Conclusion

- We propose a novel idea that uses a lightweight image compression encoder on satellites and leverages satellite images' text-to-image pairing nature for compensation when decompressing.
- We propose a novel compensation model based on stable diffusion to compensate image details when decompressing with the unique sensor data of satellite images as descriptions.
- We consider two datasets under satellite image transmission scenarios, taking into account the typical satellite image transmission tasks like tile scenes, and show SOTA performace.

Thank you for listening!