

CodecSplat: Ultra-Compact Latent Coding for Feed-Forward 3D Gaussian Splatting

Pengpeng Yu^{1,2}, Runqing Jiang¹, Qi Zhang^{2,3}, Dingquan Li², Jing Wang², Yulan Guo^{1,2}

¹ Sun Yat-sen University ² Pengcheng Laboratory ³ Peking University

Explore the code and demo

github.com/pengpeng-yu/CodecSplat



Compress the features that generate Gaussians, not the Gaussians themselves.

01 Why compress latent features?

Feed-forward models reconstruct scenes quickly, but their dense Gaussian outputs require substantial storage.

Post-hoc Images → 2D features → 3D Gaussians → Codec → 3D Gaussians

Ours Images → 2D features → **Codec** → 3D Gaussians

CodecSplat stores compressed features and camera parameters in one bitstream. The decoder reconstructs the Gaussians without the input images.

03 One model supports multiple rates

We evaluate DL3DV with 8 views at 256×448 and RE10K with 2 views at 256×256.

Method / setting	DL3DV		RE10K	
	KiB / scene	PSNR (dB)	KiB / scene	PSNR (dB)
DepthSplat + PNG (16/8)	11,942.05	24.94	1,678.17	27.28
ReSplat + PNG (8/4)	447.16	19.79	890.08	22.38
ReSplat + SOGS (8/4)	604.84	23.46	1,119.22	26.25
ReSplat + SOGS (16/8)	879.45	28.36	1,841.91	29.36
ReSplat + FCGS	1,425.22	28.72	2,526.32	29.16
CodecSplat ($\lambda = 16$)	20.00	23.56	3.37	24.76
CodecSplat ($\lambda = 64$)	41.52	25.24	5.58	26.18
CodecSplat ($\lambda = 1024$)	107.77	26.36	12.51	27.05

PNG/SOGS settings specify the bits for means and attributes. PNG uses 8,192 SH clusters; FCGS uses $\lambda = 0.0016$.

CodecSplat trades some peak quality for substantially smaller bitstreams and faster encoding.

Encode / decode time (s)	DL3DV	RE10K
DepthSplat + PNG (16/8)	18.277 / 0.487	2.778 / 0.075
ReSplat + SOGS (16/8)	16.114 / 0.024	60.284 / 0.062
ReSplat + FCGS	0.762 / 0.969	1.062 / 1.271
CodecSplat (rate average)	0.267 / 0.273	0.062 / 0.059

04 Why use refinement and learned coding?

Pre- and post-refinement reduce BD-rate on DL3DV by 13.67% and 18.08%, respectively, compared with removing each module.

DL3DV variant	Size	PSNR (dB)
Without pre-refiner	130.13 KiB	26.29
Without post-refiner	131.65 KiB	26.28
FP16 + Deflate	81.08 MiB	26.22
Full CodecSplat	107.77 KiB	26.36

Learned variants use $\lambda = 1024$. Deflate stores FP16 features and predicted depth.

05 Can the codec transfer to other FF3DGS backbones?

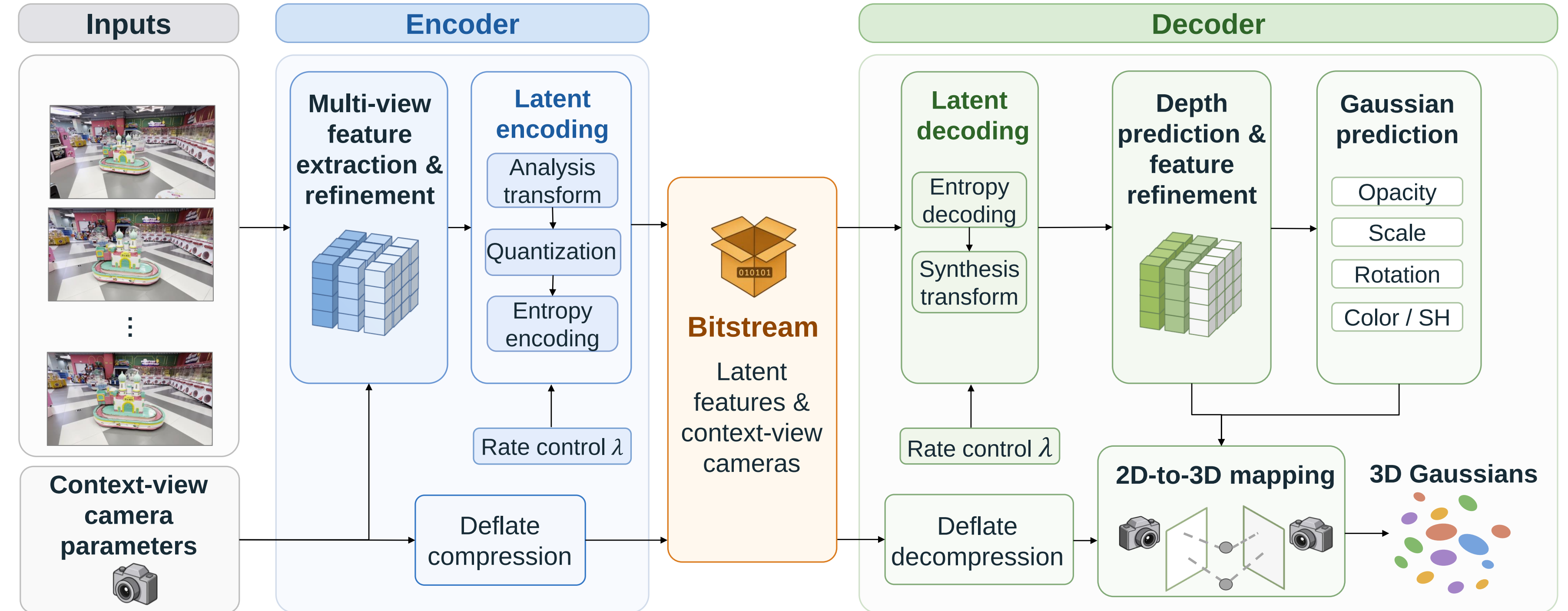
We train ZPressor + CodecSplat with 12 views and 6 clusters, then test the same model with 12-36 input views.

Views	KiB / scene	PSNR (dB)
12	85.65	24.66
16	86.02	24.60
24	86.25	24.59
36	86.30	24.16

At $\lambda = 1024$, PSNR is 0.28-0.36 dB higher than the published ZPressor results.

02 Compress features before Gaussian construction

We compress multi-view features, then reconstruct Gaussians from the decoded features and cameras.



We first train Gaussian generation, then train the codec with rendering, depth, and rate losses.

06 Compact scenes retain visual detail

Target view	DepthSplat + PNG	ReSplat + FCGS	CodecSplat
Reference	22.06 dB 12.09 MiB	24.60 dB 1.40 MiB	22.82 dB 93.16 KiB
Reference	20.52 dB 11.97 MiB	24.47 dB 1.31 MiB	22.45 dB 120.23 KiB
Reference	24.88 dB 11.54 MiB	31.73 dB 1.50 MiB	28.52 dB 87.29 KiB

Labels show target-view PSNR and compressed scene size; CodecSplat uses $\lambda = 1024$.