Bound-Conditioned Latent Inference for Progressive Image Compression

Jaeseok Jang[1†], Seungmin Jeon[1†], Kwang Pyo Choi[2], Chang-Su Kim[1]

[1]Korea University [2]Samsung Electronics [†]Equal contribution

Background

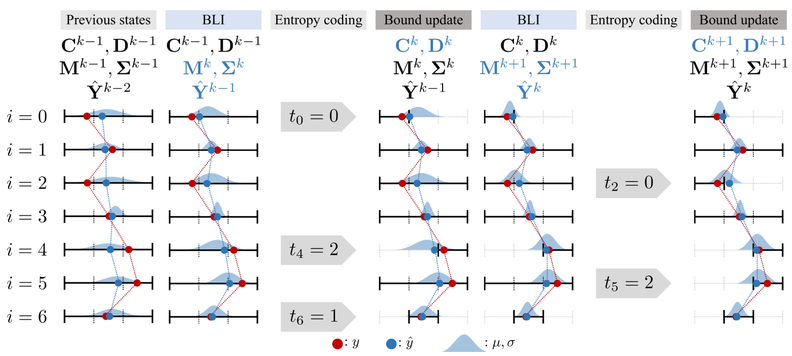
Progressive image compression and trit-plane coding

A progressive codec reconstructs an image from any prefix of a single bitstream, so one encoding serves every quality level (adaptive streaming, bandwidth-limited transmission).

- **Trit-plane coding** (DPICT [1]): each latent element is written in ternary digits (trits) and transmitted plane by plane, from the most significant trit down.
- CTC [2] adds context models that refine trit probabilities and partial reconstructions.
- **Limitation:** decoded trits are used only as discrete symbols. The feasible interval they impose on each continuous latent is not exploited for in-plane latent inference.

Contributions

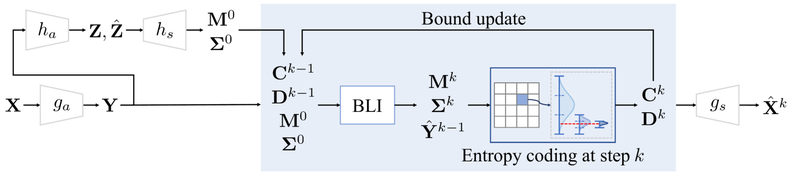
- We reformulate progressive trit-plane coding as **sequential latent inference under shrinking interval constraints**.
- We propose **BLI**, a learned module that jointly refines the entropy-model parameters and the latent estimate from bound and hyperprior information, with flexible in-plane refinement steps.
- State-of-the-art R-D among progressive codecs: BD-rate **−17.78%** vs. DPICT and **−7.96%** vs. CTC on Kodak, with 5.3× fewer parameters and lower latency than CTC.



Each decoded trit keeps one third of the current interval, so the feasible range of every latent shrinks. BLI re-estimates the Gaussian (blue) and the estimate $\hat{y}$ (blue dot) within the tighter bounds; red dots are the true latents y.

Bound-Conditioned Latent Inference

Step-wise inference within the feasible intervals



At step k, BLI maps the bounds $(\mathbf{C}^{k-1}, \mathbf{D}^{k-1})$ and the hyperprior $(\mathbf{M}^0, \mathbf{\Sigma}^0)$ to $(\mathbf{M}^k, \mathbf{\Sigma}^k)$ for entropy coding and $\hat{\mathbf{Y}}^{k-1}$ for reconstruction; decoded trits then tighten the bounds to $(\mathbf{C}^k, \mathbf{D}^k)$.

Trits decoded so far confine each latent element y_i to an interval $[\mathbf{L}_i, \mathbf{U}_i]$. BLI is conditioned on its center $\mathbf{C} = (\mathbf{L} + \mathbf{U})/2$ and width $\mathbf{D} = \mathbf{U} - \mathbf{L}$.

Entropy modeling. The interval is split into three sub-intervals, one per trit value, and the trit t_i is coded with the Gaussian probability mass of its sub-interval:

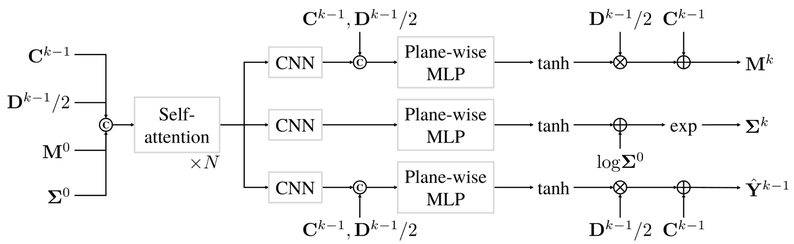
$$\mathcal{L}_R^k = -\sum_i \log_2 P(y_i \in \mathcal{I}_{i,t_i}^{k-1} \mid \mathbf{M}_i^k, \mathbf{\Sigma}_i^k)$$

Reconstruction. BLI predicts a residual around the interval center, trained with the distortion of the decoded image:

$$\hat{\mathbf{Y}}^{k-1} = \mathbf{C}^{k-1} + \Delta\hat{\mathbf{Y}}^{k-1}, \mathcal{L}_D^k = \|\mathbf{X} - g_s(\hat{\mathbf{Y}}^k)\|_2^2$$

BLI Module

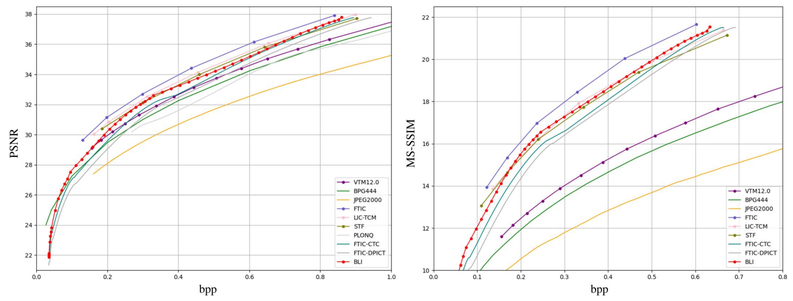
One shared network; outputs always stay inside the feasible interval



- Swin self-attention blocks model spatial dependencies among partially decoded latents; plane-wise MLP heads adapt to each trit-plane level.
- $\mathbf{M}^k$ and $\hat{\mathbf{Y}}^{k-1}$ pass through tanh and are rescaled by $\mathbf{D}^{k-1}/2$ around $\mathbf{C}^{k-1}$, which keeps them consistent with the bounds.
- Each trit-plane can be split into several refinement steps (4 per plane with random trit assignment in training; $\mathcal{K}_{\text{end}} = \{4, 8, 14, 30, 54\}$ in evaluation).

Experimental Results

Kodak · CLIC 2021 · JPEG-AI; DPICT and CTC rebuilt on FTIC [3]



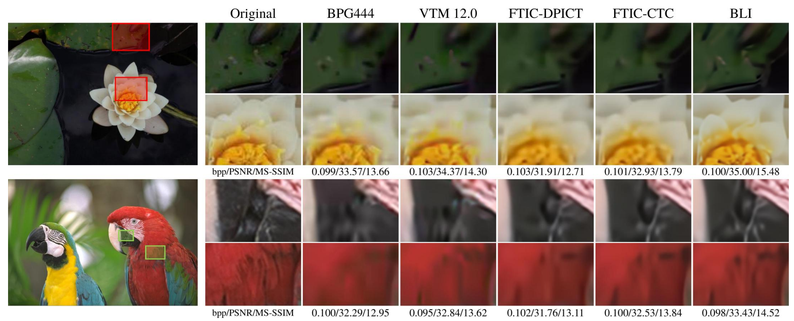
R-D curves on Kodak: PSNR (left) and MS-SSIM in dB (right). All BLI points come from one progressive bitstream, while STF, LIC-TCM, and FTIC are trained per rate.

Table 1. PSNR BD-rates (%) relative to FTIC-DPICT (lower is better)

	Method	Kodak	CLIC	JPEG-AI
Traditional	JPEG2000	47.54	58.41	55.73
	BPG444	−7.92	0.21	0.74
	VTM 12.0	−2.14	−2.17	−0.85
Fixed-rate	STF	−14.85	−12.87	−11.69
	LIC-TCM	−18.69	−19.37	−18.18
	FTIC	−26.37	−23.33	−20.49
Progressive	PLONQ	14.83	-	21.29
	FTIC-CTC	−9.82	−11.52	−11.53
	BLI	**−17.78**	**−17.12**	**−17.52**

Table 2. Parameter counts and latencies

	Method	# Params. (M)	Latency (s)
Fixed-rate	LIC-TCM	76.57	6.53
	FTIC	71.01	215.73
Progressive	FTIC-DPICT	40.53	5.85
	FTIC-CTC	1,032.65	6.62
	BLI	**196.12**	**4.77**



Qualitative comparison at nearly 0.1 bpp (bpp / PSNR / MS-SSIM). BLI better preserves petal and feather textures.

References

[1] J.-H. Lee et al. DPICT: Deep progressive image compression using trit-planes. CVPR 2022.
[2] S. Jeon et al. Context-based trit-plane coding for progressive image compression. CVPR 2023.
[3] H. Li et al. Frequency-aware transformer for learned image compression. ICLR 2024.