Generalized Nested Latent Variable Models For Lossy Coding Applied To Wind Turbine Scenarios

ICIP abu dhabi 2024
IEEE International Conference on Image Processing
27 - 30 October 2024 | Abu Dhabi, UAE

Raül Pérez-Gonzalo, Andreas Espersen and Antonio Agudo

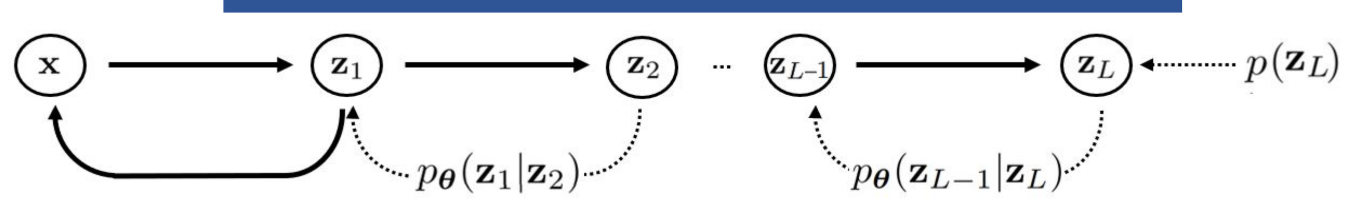
Institut de Robòtica i Informàtica Industrial, CSIC-UPC, Spain
Wind Power LAB, Copenhagen, Denmark

Institut de Robòtica i Informàtica Industrial
CSIC
CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS
UPC

Abstract

Rate-distortion optimization seeks to minimize the compromise between compression rate and reconstructed image quality. A successful approach introduces a deep hyperprior in a 2-level nested latent variable model. This work extends this concept to an L-level nested generative model with a Markov chain structure. We show as L increases, a trainable prior becomes detrimental and explore shared dimensionality across latent variables to enhance compression. This framework generalizes autoregressive coders, outperforming the hyperprior model and achieving state-of-the-art performance with lower computational cost.

Nested Latent Variable Models



We aim to learn a probabilistic model $p_\theta(x)$ of our observed data x to successfully apply an entropy coder capable of compressing them, where θ denotes the model parameters. To address this learning problem, fully-observed models are iteratively marginalized over nested latent variables $z_1,\dots,z_L$

$$p_\theta(\mathbf{x}) = \int p_\theta(\mathbf{x}|\mathbf{z}_1)p_\theta(\mathbf{z}_{1:L-1}|\mathbf{z}_L)p(\mathbf{z}_L)\,d\mathbf{z}_{1:L}$$

where we have defined for conciseness:

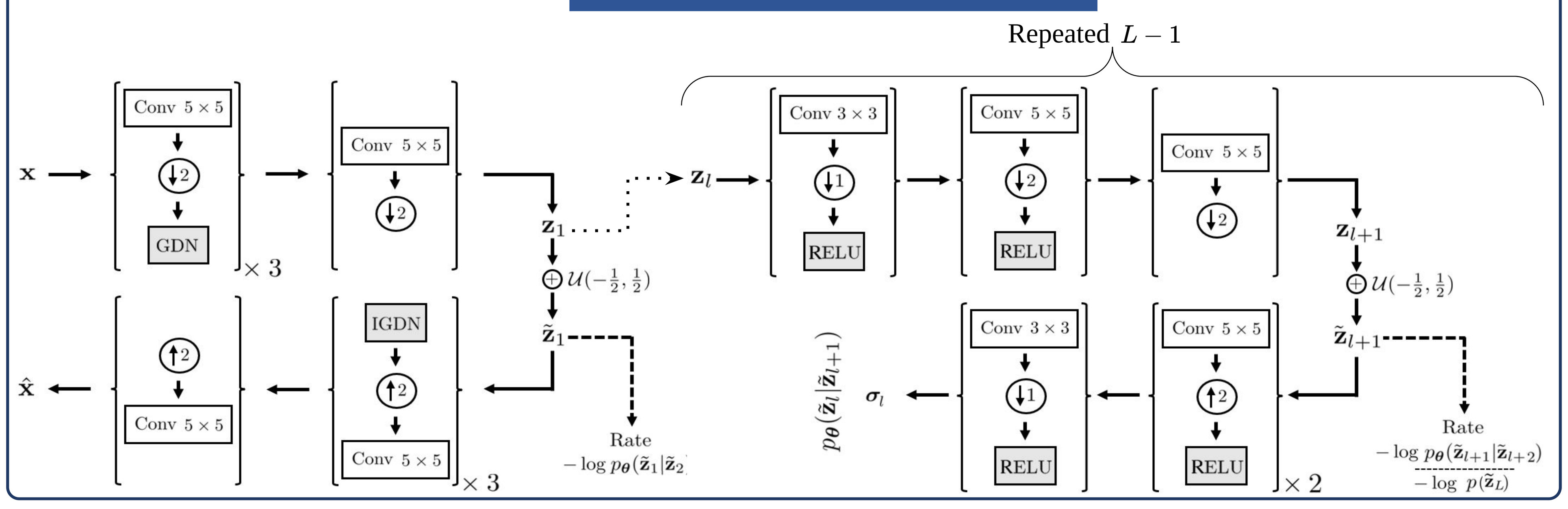
$$p_\theta(\mathbf{z}_{1:L-1}|\mathbf{z}_L) = \prod_{l=1}^{L-1} p_\theta(\mathbf{z}_l|\mathbf{z}_{l+1}) \text{ and } d\mathbf{z}_{1:L} = d\mathbf{z}_1 d\mathbf{z}_2 \dots d\mathbf{z}_{L-1} d\mathbf{z}_L$$

Generalized Rate-distortion trade-off

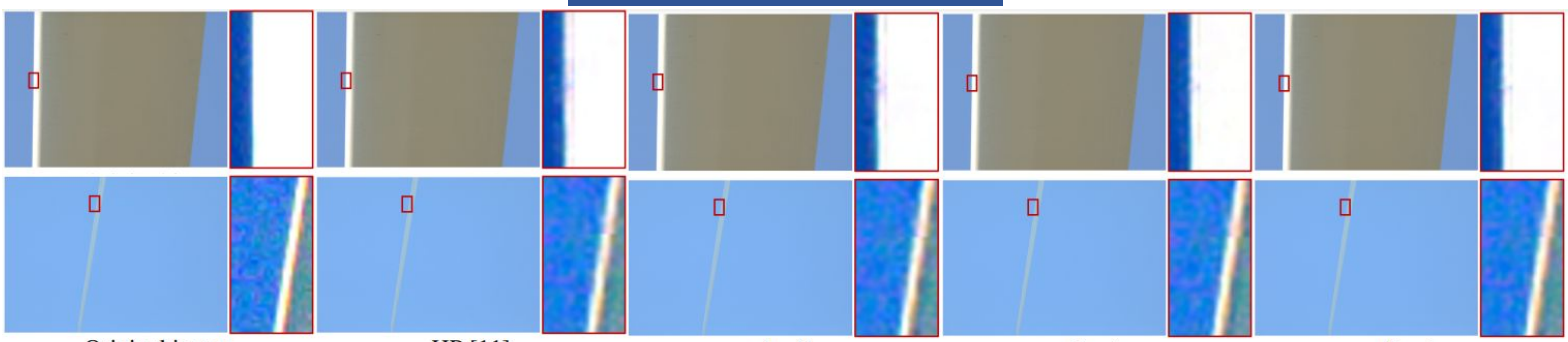
We seek to minimize the trade-off governed by $\lambda \in R$ between the compression rate and the decompression error d:

$$-\mathbb{E}_{p_\theta(\mathbf{x})}\left[\log p_\theta(\mathbf{z}_{1:L-1}|\mathbf{z}_L) + \log p(\mathbf{z}_L)\right] + \lambda\,\mathbb{E}_{p_\theta(\mathbf{x})}\left[d(\mathbf{x},\hat{\mathbf{x}})\right]$$

Model Architecture



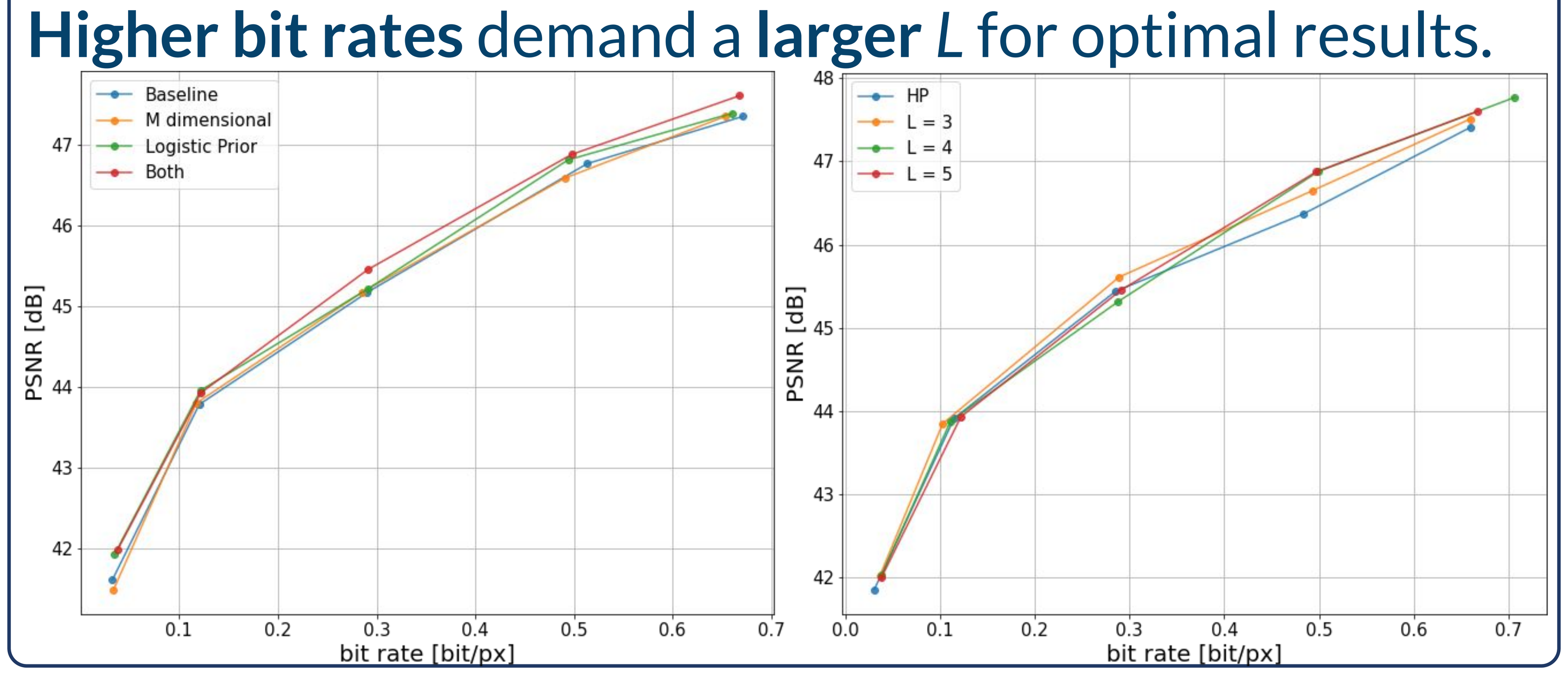
Visual Comparison: same bit rate



Original image | HP [11] | $L=3$ | $L=4$ | $L=5$

Ablation Study

A **prior entropy network** is no longer required.
Higher bit rates demand a **larger** L for optimal results.



Autoregressive (AR) Universal Approximators

Let $x^{(t)}$ denote the t-th pixel of $\mathbf{x}$, AR models leverage the previously decoded pixels $x^{(t-1)}$, $x^{(t-2)}$,..., $x^{(t-2)}$ to enhance the coding performance of the next pixel

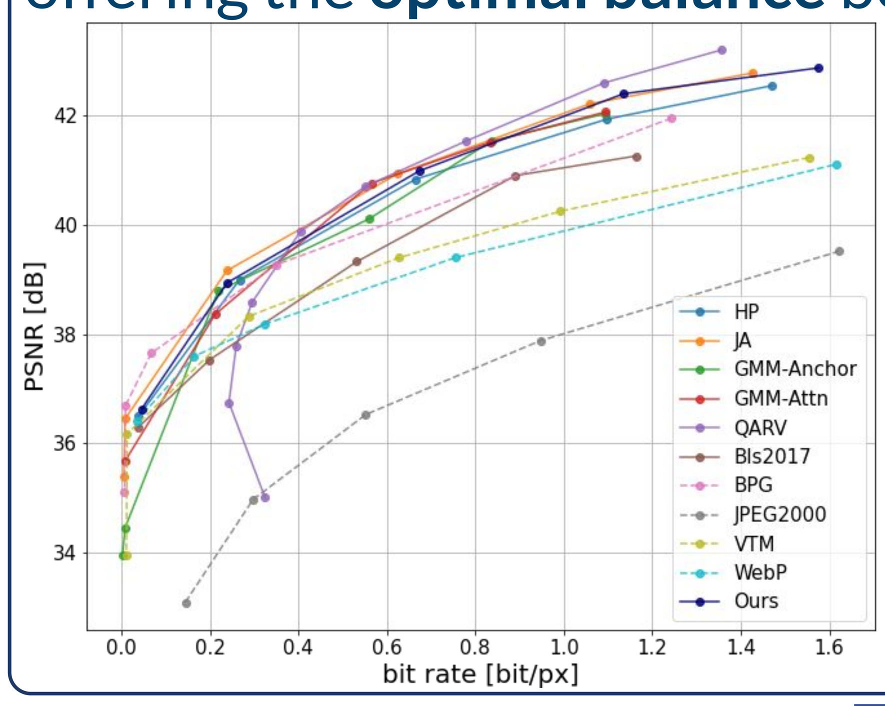
$$p_\theta(\mathbf{x}) = \prod p_\theta(x^{(t)}|x^{(t-1)},\dots,x^{(1)})$$

WLOG, we assume each latent variable z_l only has a single component. Nested latent variable models can approximate AR models by decoding each pixel component through an additional latent variable: the t-th pixel component $x^{(t)}$ is decoded by z_{L-t+1} through

$$p_\theta(\mathbf{x}) = \prod p(z_{L-t+1}|z_{L-t+2},\dots,z_L)$$

SOTA Comparison

Our model delivers **robust, high-performance coding** with a computational cost that scales linearly with nested latent variables, offering the **optimal balance** between efficiency and **processing time**.



Coding scheme	Compression Time: Minimum time [s]	Maximum time [s]	Mean time [s]	Slope growth [s/bit]
HP [11]	5.76	6.17	5.96	0.25
JA [15]	200.68	204.19	203.01	1.96
GMM-Anchor [22]	227.91	235.96	233.04	5.51
GMM-Attn [22]	247.97	258.12	252.81	6.78
Bls2017 [10]	0.27	0.35	0.30	0.07
QARV [14]	371.08	374.68	373.28	2.21
BPG [4]	2.47	8.80	4.75	3.25
JPEG2000 [1]	7.02	7.03	7.02	0.01
VTM [5]	102.57	9943.02	4584.58	5309.94
WebP [2]	2.49	9.40	4.17	3.59
Ours	9.10	13.12	10.73	2.71

Conclusions

We introduce a versatile L-level nested latent model that captures the intricate dependencies among latent variables with greater fidelity and marked compression improvement. By selecting the optimal layer depth depending on the rate-distortion trade-off, these generalized models surpass the hyperprior performance without a trainable prior and successfully approximate autoregressive models, accomplishing state-of-the-art results while reducing the computational cost.

Supported by Innovation Fund Denmark under 2021 ID1044-0044A and by MoHuCo PID2020- 120049RB-I00; MCIN/AEI/10.13039/501100011033.

WIND POWER LAB
Global Blade Optimisation