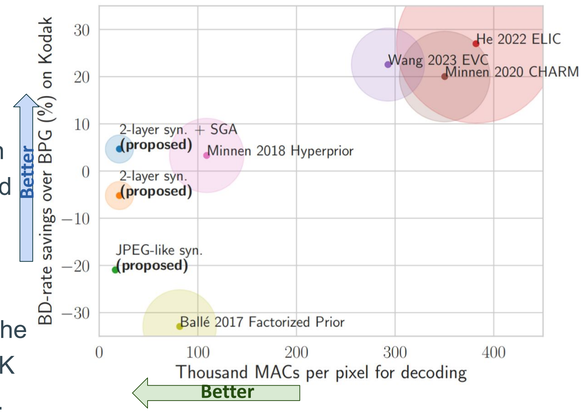
Computationally-Efficient Neural Image Compression with Shallow Decoders

Yibo Yang, Stephan Mandt

Paper: https://arxiv.org/abs/2304.06244 **Data & code:** https://github.com/mandt-lab/shallow-ntc

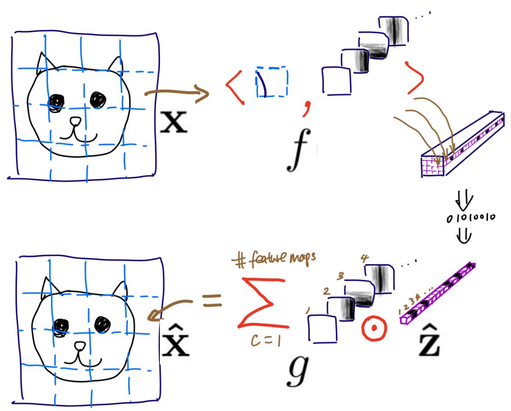
Overview:

- Neural image compression methods have beaten classical codecs in rate-distortion performance, but require much higher computation.
- While existing autoencoder architectures use symmetrical encoders and decoders, many applications (e.g., mobile, streaming) have an asymmetrically lower *decoding* computation budge than for *encoding*.
- Motivated by this, we reduce the decoding complexity of neural compression methods using 1. extremely shallow synthesis transforms inspired by JPEG; and 2. more powerful and expensive encoding methods.
- Our results establish a new frontier in the trade-off between rate-distortion and decoding complexity for neural image compression, in the regime of sub-50K FLOPs per pixel .



Background: transform coding

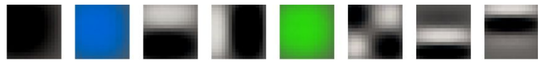
- Traditional image compression (e.g., JPEG) works by block-wise transform coding, via a pair of (analysis, synthesis) transform, $(\boldsymbol{f}, \boldsymbol{g})$.
  - An image block $\boldsymbol{x}$ is encoded into a vector of (latent) coefficients $\boldsymbol{z}$, entropy coded, and decoded into a reconstructed block $\boldsymbol{x}$;
  - The encoding/decoding transforms are linear maps.



- *Neural* image compression methods implement $(\boldsymbol{f}, \boldsymbol{g})$ with deep convolutional neural nets (CNNs) instead of orthogonal matrices.

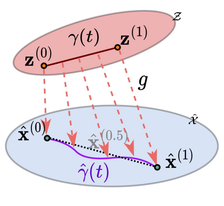
Neural image compression approximates linear transform coding

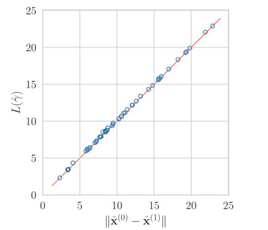
- By decoding one-hot "basis" tensors in the latent space, previous work [Duan et al., 2022] observed that the synthesis transform can be interpreted as implementing classical transform basis functions:



16x16 images decoded from "bases" latent tensors using a hyperprior architecture. Duan et al., 2022.

- Here, we further examine the approximate affine behavior of learned neural synthesis transforms, showing that they map straight paths in the latent space to approximately straight paths in the *pixel space*:
- This forms a stark contrast to the typical behavior of the decoder network in generative modeling, where varying the latents have more global effects on the decoded output.





Scatter of curve lengths of latent interpolation connecting two random images, v.s. the Euclidean distance between the two images.

Example decoded interpolation in the latent (coefficient) space.

Example interpolation in the pixel space. The result is nearly the same.

Proposed: asymmetrically-powered compression

- Given the close connection to transform coding, we conjecture that a deep convolutional synthesis network might be replaceable by a linear one similar to the DCT in JPEG (but learned end-to-end).
- Pros: significant savings in computation complexity, straightforward to implement with a single transposed convolution layer. Con: potentially a large drop in R-D (rate-distortion) performance.
- Solution:
  - We modify the JPEG synthesis to use larger/overlapping blocks, and add a small amount of non-linearity for more expressiveness;
  - We make the encoding procedure more powerful, e.g., with a bigger network and potentially iterative inference [Yang et al. 2020].
- We formally justify our approach by deriving a novel decomposition of the asymptotic R-D loss of neural lossy compression, quantifying the effect of different transform/modeling choices in isolation.

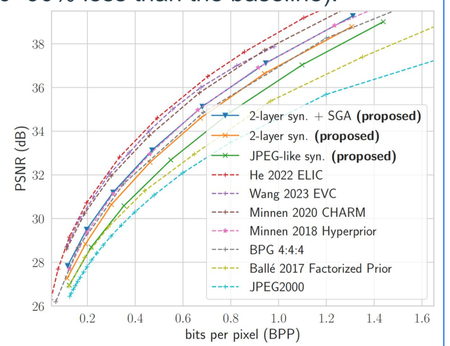
Experiments

Setup:

- We start with the mean-scale hyperprior [Minnen et al., 2018] as our base architecture.
- We upgrade its CNN analysis transform to incorporate attention [He et al., 2022] and possibly followed by iterative inference with SGA [Yang et al. 2020]. We upgrade the CNN synthesis transform by either:
  - a (single-layer) JPEG-like block-wise linear transform; or
  - a two-layer shallow transform with a single non-linearity.

Findings:

- WIth a large enough kernel size, a JPEG-like synthesis can largely match the R-D performance of a much more expensive CNN synthesis, with reconstruction quality measured in PSNR.
- Compared against other SOTA computationally-efficient neural image compression methods, including ELIC [He et al., 2022] and EVC [Wang et al., 2023], our proposed asymmetric architectures achieve a new Pareto frontier for rate-distortion-complexity.
- Specifically, our two-layer synthesis transform + SGA performs on par with BPG and the base architecture [Minnen et al., 2018] at less than 50K decoding FLOPs/pixel (80~90% less than the baseline).
  - The JPEG-like synthesis can still encounter blocking artifacts at low bitrates, despite using a large kernel.
  - The non-linearity is still important for the shallow synthesis to achieve perceptually better reconstructions.



Limitations and future work

- While we kept the hyperprior fixed for simplicity, a more lightweight entropy model can reduce the decoding complexity even further.
- This work focused on the PSNR quality metric; high perceptual quality at low computation complexity / bitrate remains an open question.

References

[Duan et al., 2022]. Zhihao Duan, Ming Lu, Zhan Ma, and Fengqing Zhu. Opening the black box of learned image coders. In 2022 Picture Coding Symposium (PCS)
[Minnen et al., 2018] D. Minnen, J. Ballé, and G. D. Toderici. Joint Autoregressive and Hierarchical Priors for Learned Image Compression. In Advances in Neural Information Processing Systems 31. 2018.
[Yang et al. 2020] Yibo Yang, Robert Bamler, and Stephan Mandt. Improving inference for neural image compression. In Neural Information Processing Systems (NeurIPS), 2020.
[He et al., 2022] Dailan He, Ziming Yang, Weikun Peng, Rui Ma, Hongwei Qin, and Yan Wang. Elic: Efficient learned image compression with unevenly grouped space-channel contextual adaptive coding. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022.
[Wang et al., 2023] Guo-Hua Wang, Jiahao Li, Bin Li, and Yan Lu. EVC: Towards real-time neural image compression with mask decay. In International Conference on Learning Representations, 2022.