EVC: Towards Real-Time Neural Image Compression with Mask Decay

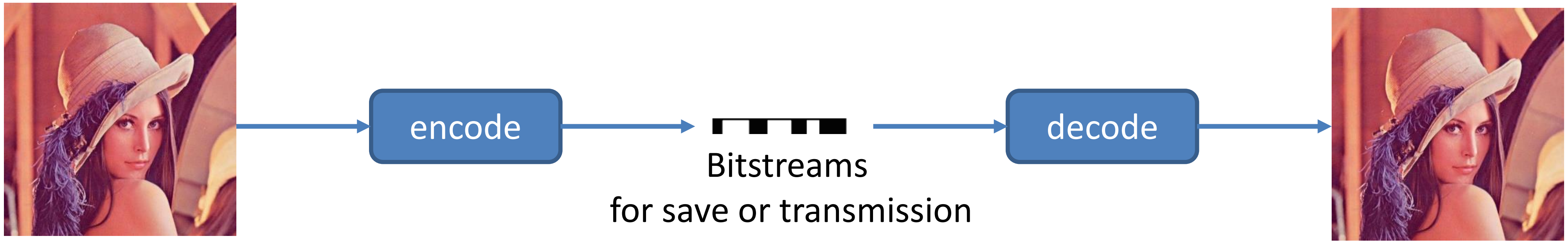
Guo-Hua Wang[1], Jiahao Li[2], Bin Li[2], Yan Lu[2]

[1]State Key Laboratory for Novel Software Technology, Nanjing University

[2]Microsoft Research Asia

Introduction

- Task: image compression

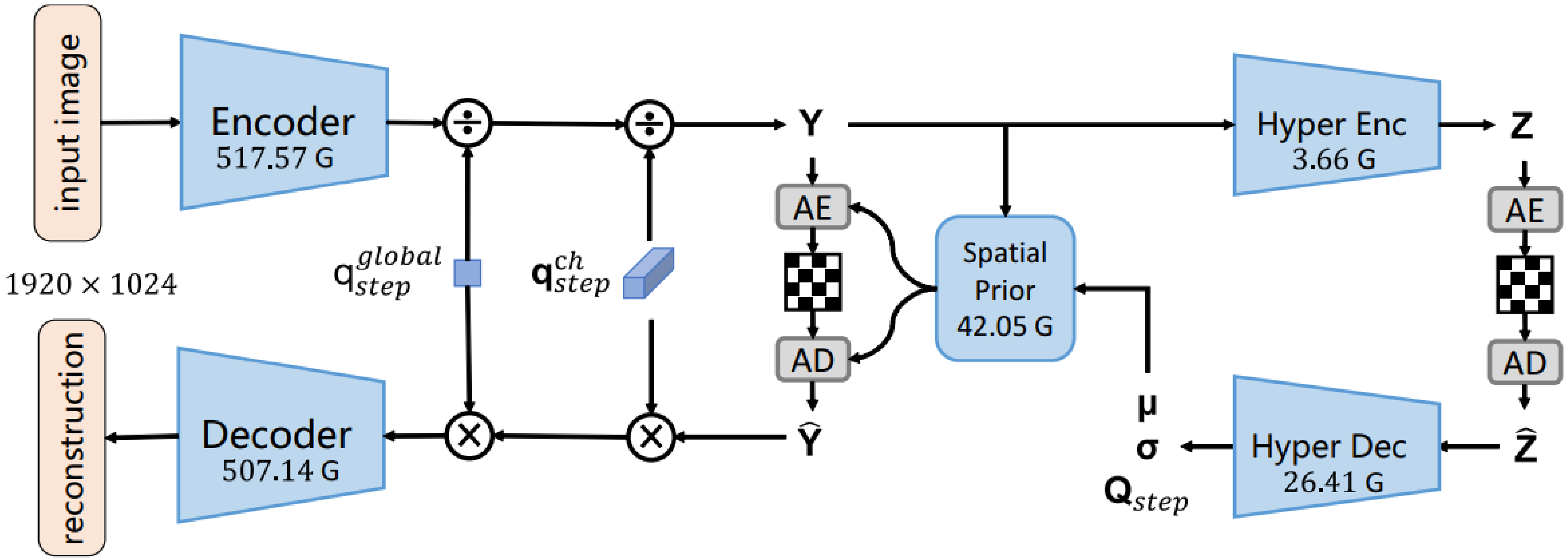


- Traditional image compression
  - JPEG, BPG, VTM
  - hand-crafted features
- Learning based image compression
  - End-to-end optimization
  - Outperform traditional methods for the rate-distortion (RD) performance
  - But it suffers from a large complexity

Our contributions

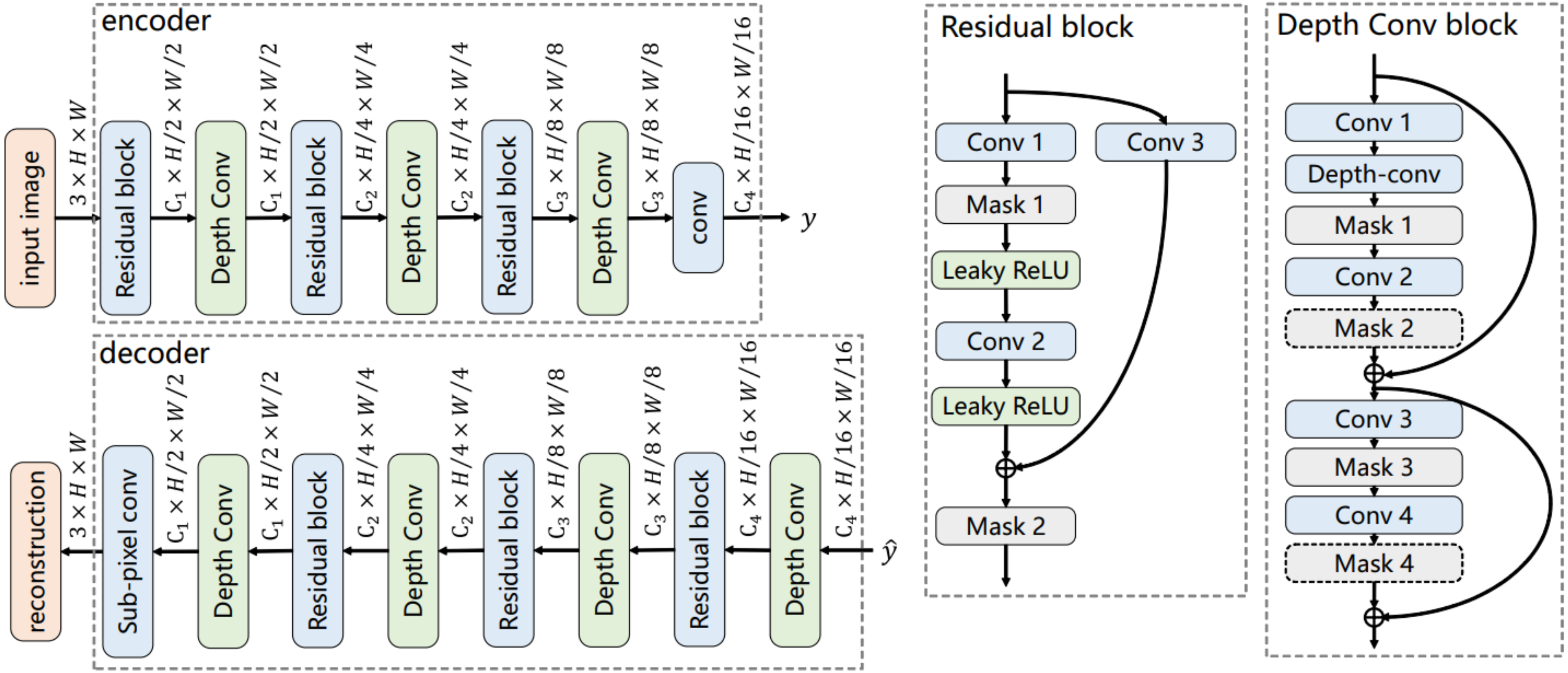
- We propose an Efficient Variable-bit-rate Codec (EVC) for image compression
  - Our Large model: 30 FPS for the 768x512 inputs
  - Our Small model: 30 FPS for the 1920x1080 inputs
  - On-par with SOTA models for the RD performance
- We propose mask decay with a novel sparsity criterion
  - Our medium and small models are improved significantly by 50% and 30%, respectively.
- We advocate the scalable encoder for neural image compression
  - With residual representation learning and mask decay, our scalable encoder achieves a superior complexity-RD trade-off

Our EVC framework



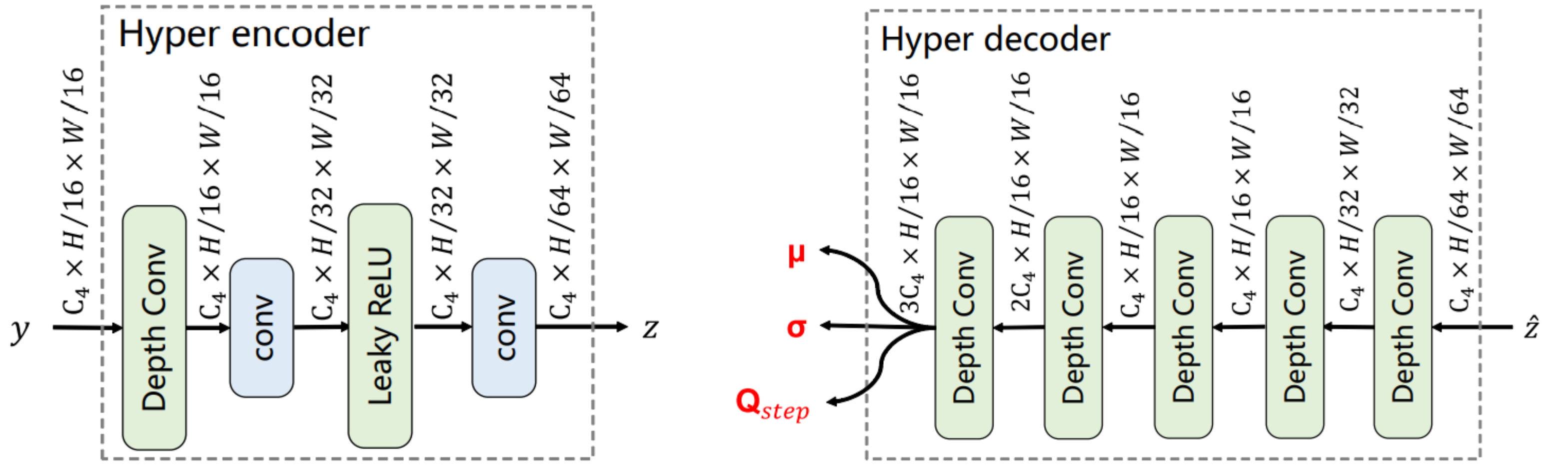
- We introduce adjustable quantization steps for variable RD trade-offs.
- Both encoder and decoder suffer from large complexities

Encoder and Decoder



Model	C_1, C_2, C_3, C_4		Encoder	Decoder	Others	Total
Large (L)	192, 192, 192, 192	#Params MACs	3.19 517.57	3.38 507.14	10.82 72.17	17.38 1096.84
Medium (M)	128, 128, 192, 192	#Params MACs	2.08 247.43	2.33 243.35	10.82 72.17	15.23 562.91
Small (S)	64, 64, 128, 192	#Params MACs	0.82 68.87	1.14 69.48	10.82 72.17	12.78 210.47

Hyperprior



(a) (b)

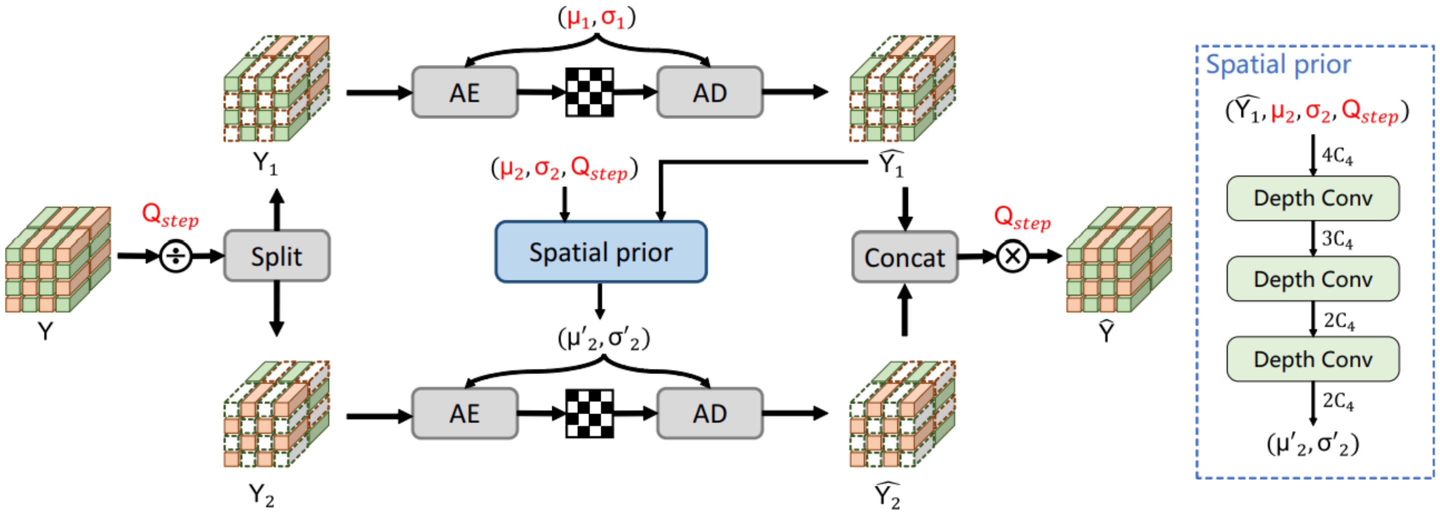
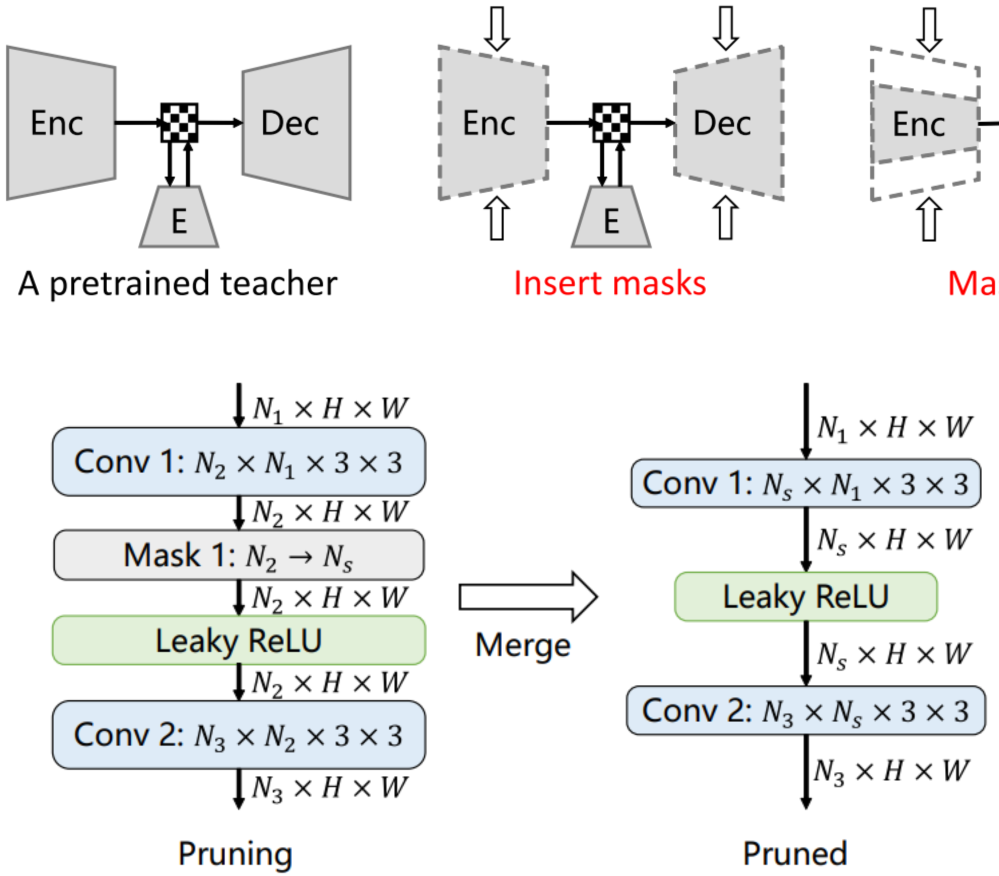
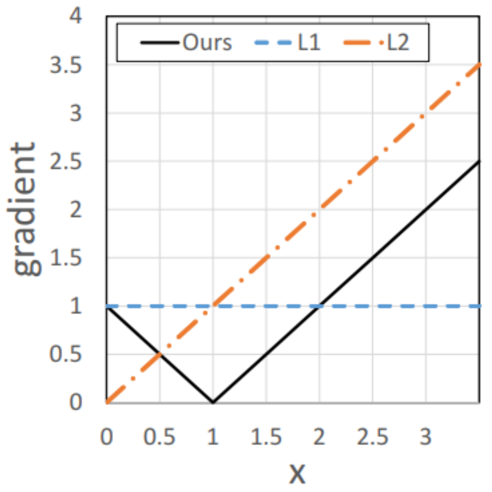


Figure 11: The structure of our dual spatial prior.

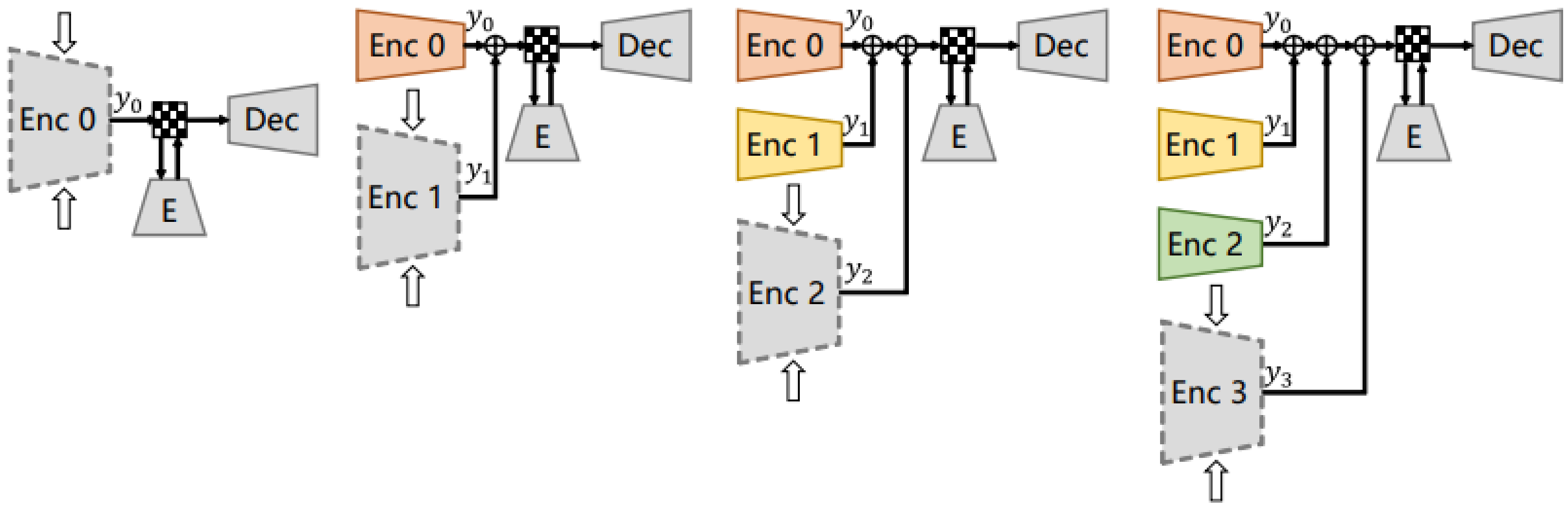
Mask decay





- The gradient of L2 norm vanishes when x approaches zero
- The gradient of L1 norm is a constant without considering its own magnitude
- Ours: $\frac{\partial \mathcal{L}_{sparse}(x)}{\partial x} = |x-1|\,, \quad \mathcal{L}_{sparse}(x) = \begin{cases} -\frac{1}{2}x^2 + x, & \text{if } 0 \leq x \leq 1\,, \\ \frac{1}{2}x^2 - x + 1, & \text{if } x > 1\,. \end{cases}$

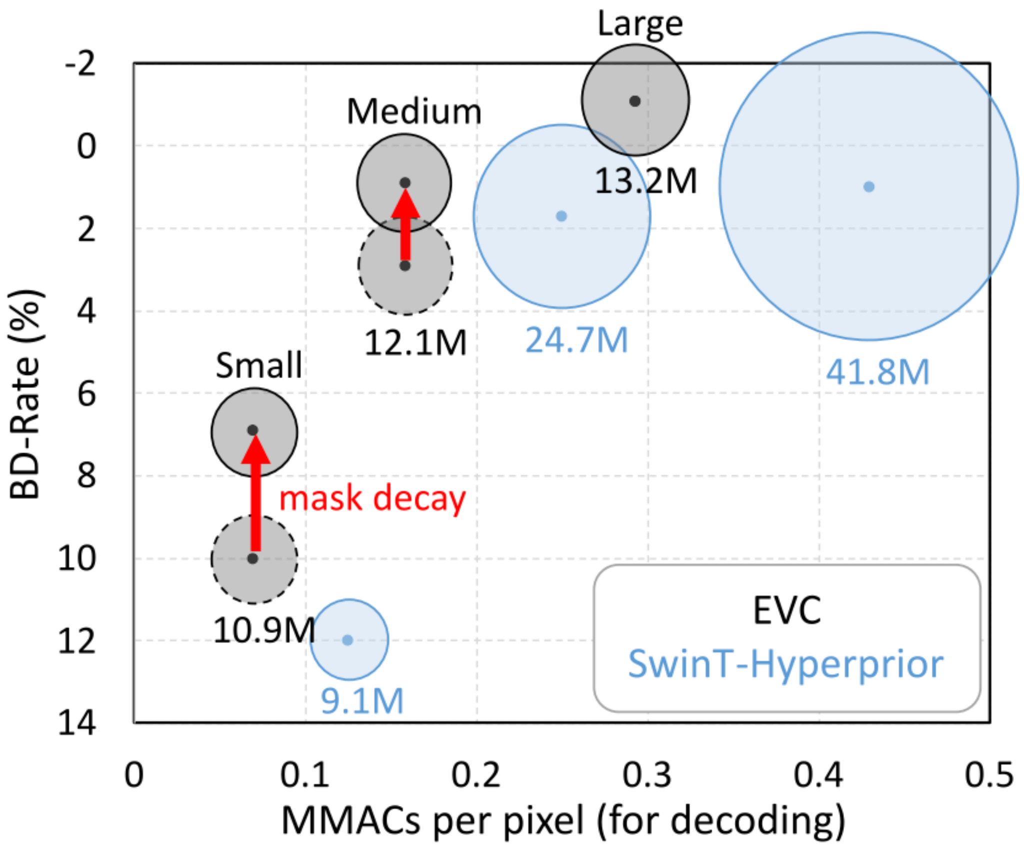
The scalable encoder

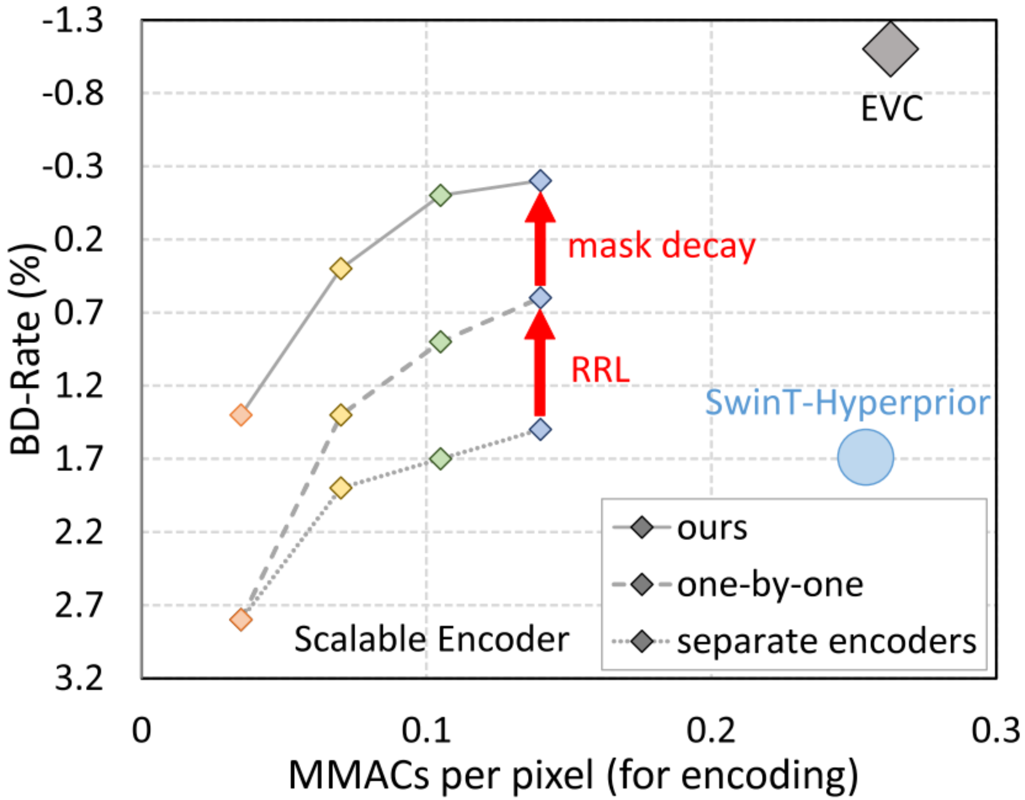


- Residual representation learning (RRL) encourages the encoder's diversity
- Both RRL and mask decay treat the teacher as a reference, which makes the training more effective

Experiments

- Mask deacy and our scalable encoder





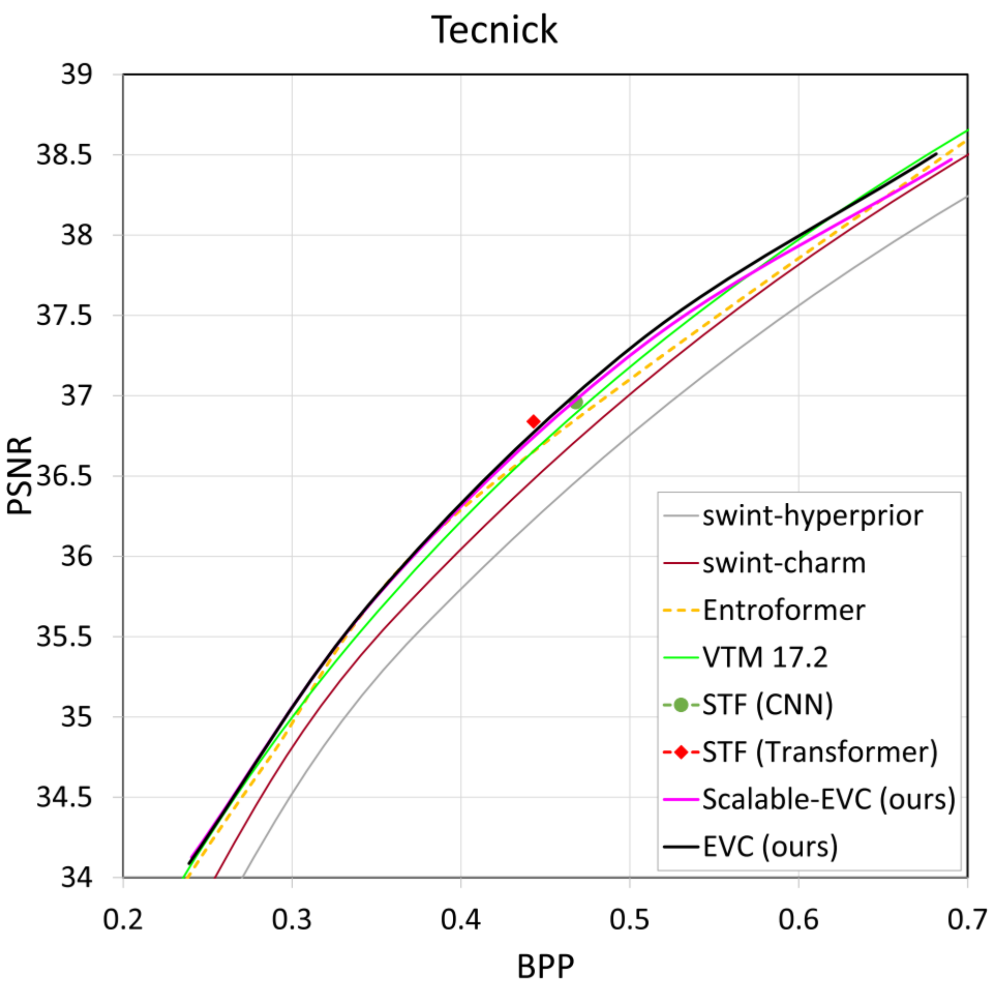
Latency

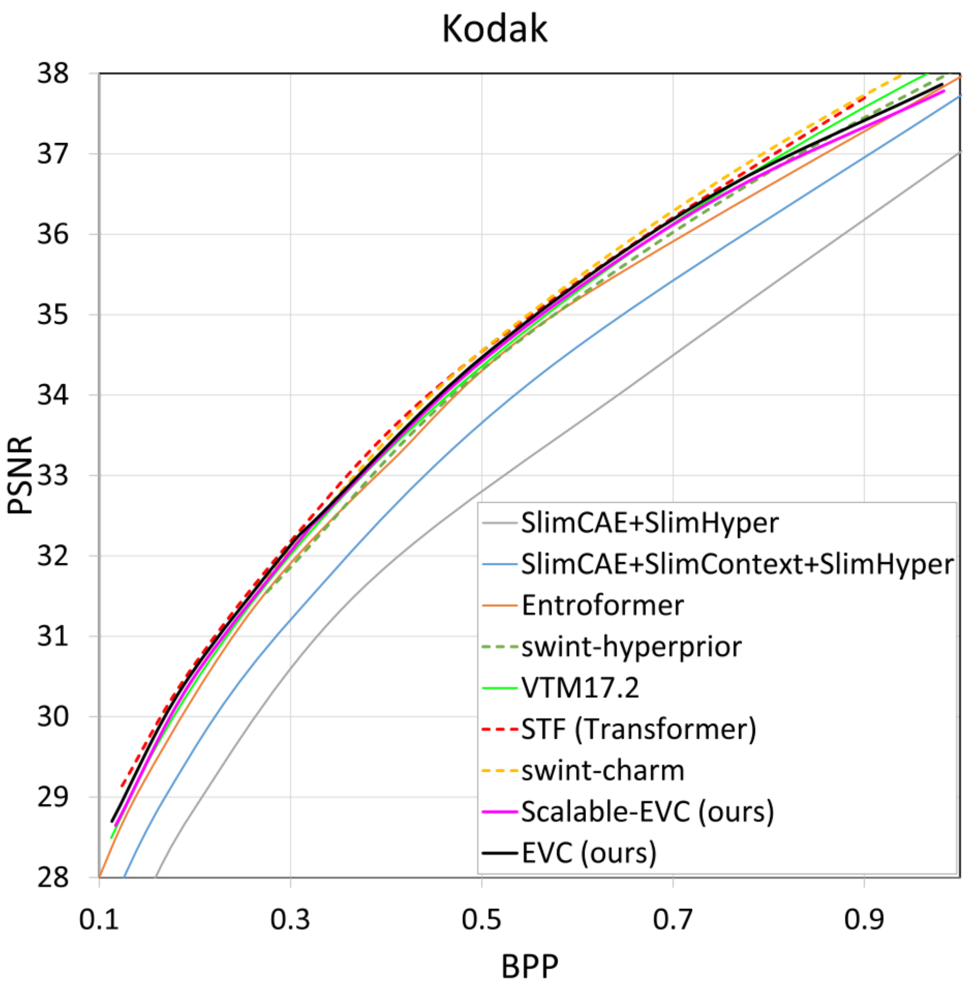
- Comparison with state-of-the-art

Resolution	GPU	Type	Entroformer	STF Transformer	STF CNN	EVC Large	EVC Medium	EVC Small
768×512	2080Ti	encoding	OM	176.3	158.5	63.0	44.7	**28.4**
		decoding	OM	202.3	210.2	41.1	**32.4**	**24.4**
	A100	encoding	816.8	115.9	96.4	**21.1**	**19.8**	**17.7**
		decoding	4361.9	143.2	118.0	**19.1**	**17.1**	**15.6**
1920×1080	2080Ti	encoding	OM	576.0	456.0	305.3	181.5	90.9
		decoding	OM	531.7	652.0	179.2	118.1	73.2
	A100	encoding	7757.4	355.6	278.1	84.2	56.3	**31.4**
		decoding	OM	354.8	281.7	60.2	46.5	**29.7**

RD Curves

- Comparison with state-of-the-art





Visualization

Figure 20: Visualization of our models' reconstruction. EVC-SS denotes our model equipped with the small encoder and the small decoder, while M and L means medium and large, respectively. Numbers in the tuple are BPP, PSNR, the encoding time (ms), and the decoding time (ms), respectively. Note that the latency is measured on a computer with 2080Ti GPU. Our models are dramatically faster than VTM.

Conclusions

- A new milestone
  - Real-Time
  - On-par with SOTA for RD performance
  - A uniform model handles variable RD trade-offs
- We proposed mask decay with a novel sparse criterion
  - Our medium and small models are improved significantly by 50% and 30%, respectively.
  - The encoder is more redundant than the decoder.
- We advocate the scalable encoder for neural image compression
  - With residual representation learning and mask decay, our scalable encoder achieves a superior complexity-RD trade-off

Thank you!

https://openreview.net/pdf?id=XUxad2Gj40n

https://github.com/microsoft/DCVC/tree/main/EVC