Idempotence and Perceptual Image Compression

Tongda Xu, Dailan He, Ziran Zhu, Yanghao Li, Lina Guo, Yuanyuan Wang, Zhe Wang, Hongwei Qin, Yan Wang, Jingjing Liu, Ya-Qin Zhang

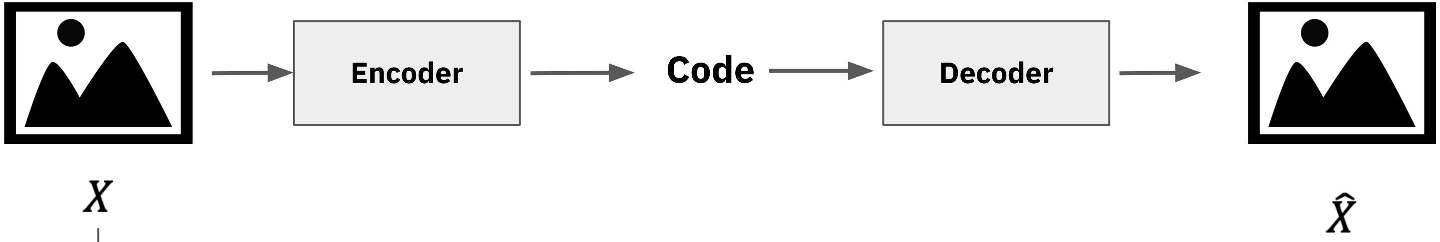
ICLR 2024 Spotlight

02. 23. 2023
EffL @POSTECH
Hagyeong Lee

https://arxiv.org/abs/2401.08920

Overview : Lossy Image Compression

- 3 Trade-off properties of lossy image compression
  => Rate-Distortion-Perception



Similar to original image => Use more bits

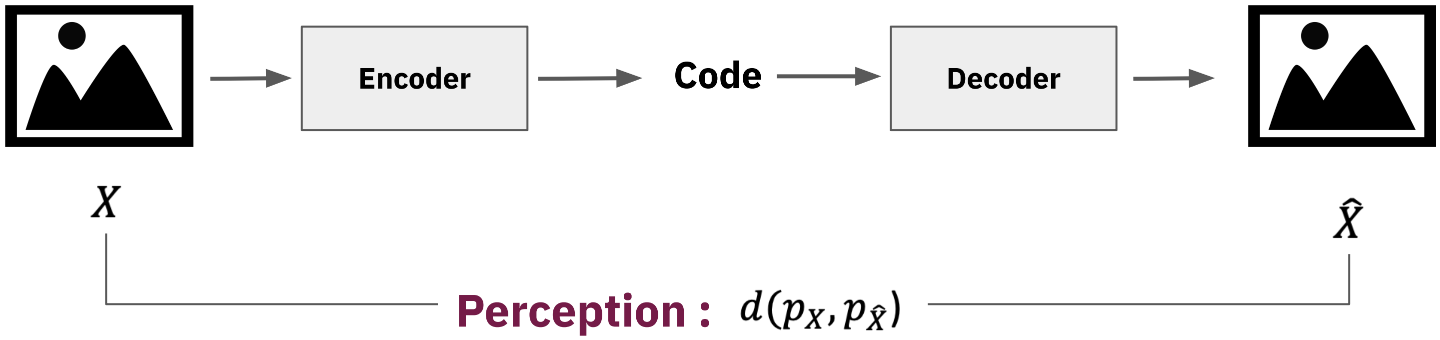
=> High rate = High PSNR

= Low Distortion (error)

= good perceptual quality

Overview : Lossy Image Compression

- 3 Trade-off properties of lossy image compression
  => Rate-Distortion-**Perception**
  :the distribution $p_{\hat{X}}$ be similar to p_X (good perceptual quality)



Idempotence of image compression

Symbol Definition.

- X: original image / $f(\cdot)$: encoder / $\mathbf{g}(\cdot)$: decoder / Y : code as $f(X)$ / $\hat{X}$ reconstruction

$$\hat{X} = g(Y)$$

If the codec is *idempotent*,

re-compression of reconstruction produces the same result with original image.

$$f(\hat{X}) = Y, \text{ or } g(f(\hat{X})) = \hat{X}$$

Idempotence of image compression

For better perceptual quality,

- Recent works(HiFiC, ILLM, CDC) train a conditional generative model to approximate the real image's posterior on the bitstream.

$$\hat{X} = g(Y) \sim p_{X|Y}, \text{ where } Y = f(X)$$

- The majority of perceptual codec achieves perfect quality by conditional generative:

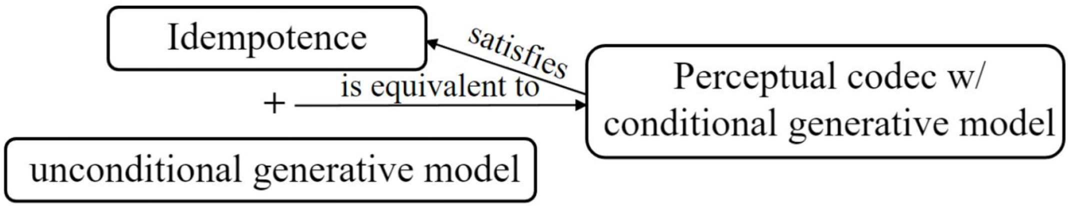
$$p_{\hat{X}} = p_X$$

← HiFiC's Qualitative Result

Idea: Perfect perceptual codec is idempotent.

To illustrate the main idea above, prove:

I. Perceptual image compression brings idempotence.
II. Idempotence brings perceptual image compression.



The relationship between idempotence and perceptual image compression.

I. Perceptual image compression brings idempotence.

$$\hat{X} = g(Y) \sim p_{X|Y} \Rightarrow f(\hat{X}) \overset{a.s.}{=} Y$$

***Proof*.**

Define the inverse image of y as : $f^{-1}[y] = \{x | f(x) = y\}$

According to definition of idempotence, need to show $\hat{X} \in f^{-1}[y]$

$Y = f(X)$ is a <u>deterministic</u> transform -> each *x* only corresponds to one *y*

Then, the likelihood of *Y* can be written as

$$p_{Y|X}(Y = y | X = x) = \begin{cases} 1, & f(x) = y \\ 0, & f(x) \neq y \end{cases}$$

I. Perceptual image compression brings idempotence.

$$\hat{X} = g(Y) \sim p_{X|Y} \Rightarrow f(\hat{X}) \overset{a.s.}{=} Y$$

***Proof*.**

Then, for all $x \notin f^{-1}[y]$, the joint distribution is :

$$p_{XY}(X = x \,|\, Y = y) = p_X(X = x) p_{Y\,|\,X}(Y = y \,|\, X = x) = 0$$

Thus the posterior is :

$$p_{X\,|\,Y}(X = x \,|\, Y = y) \; = \; 0$$

$$\Pr(\hat{X} \notin f^{-1}[y]) = 0$$

$$\Pr(\hat{X} \in f^{-1}[y]) = 1$$

$$f(\hat{X}) \overset{a.s.}{=} Y$$

II. Idempotence brings perceptual image compression.

$$\underbrace{\hat{X} \sim p_X,}\ s.t.\ \underbrace{f(\hat{X}) = Y} \Rightarrow \underbrace{\hat{X} \sim p_{X|Y}}$$

- The authors want to show sampling from unconditional generative model with idempotence constraint <u>is equivalent to</u> sampling from posterior

II. Idempotence brings perceptual image compression.

$$\hat{X} \sim p_X, \text{ s.t. } f(\hat{X}) = Y \Rightarrow \hat{X} \sim p_{X|Y}$$

***Proof*.**

(Similar to the proof of I)

Define the inverse image of y as : $f^{-1}[y] = \{x | f(x) = y\}$

$Y = f(X)$ is a deterministic transform -> the likelihood of Y can be written as

$$p_{Y|X}(Y = y | X = x) = \begin{cases} 1, & x \in f^{-1}[y], \\ 0, & x \notin f^{-1}[y]. \end{cases}$$

Then by Bayesian rule, for each $(x, y) \in \mathcal{X} \times \mathcal{Y}$,

$$p_{X|Y}(X = x | Y = y) \propto p_{Y|X}(Y = y | X = x) p_X(X = x)$$

$$\propto \begin{cases} 1 \times p_X(X = x) = p_X(X = x), & x \in f^{-1}[y], \\ 0 \times p_X(X = x) = 0, & x \notin f^{-1}[y]. \end{cases}$$

II. Idempotence brings perceptual image compression.

Equivalent to sampling from p_X with the idempotence constraint $x \in f^{-1}[y]$

Therefore, sampling from posterior :

$$\hat{X} \sim p_X, \text{ s.t. } f(\hat{X}) = Y$$

💡 Thinking different: inversion

Rewrite the left side of above Eq. as :

$$\hat{X} \sim p_X, \text{ s.t. } f(\hat{X}) = Y$$

$$\Downarrow$$

$$\min ||f(\hat{X}) - Y||^2, \text{ s.t. } \hat{X} \sim p_X$$

=> same as generative model inversion form of super-resolution

💡 Thinking different: inversion

$$\min ||f(\overset{\text{super-resolved}}{\hat{X}}) - \overset{\text{down-sampled}}{Y}||^2, \text{ s.t. } \hat{X} \sim p_X$$

In inversion, $f(\cdot)$ is down-sampling function.

In compression, the authors think $f(\cdot)$ is the encoder,

and Y is the bitstream.

Encode and Decode procedure

1. The sender(encoder) sample image from source $X \sim p_X$
2. The sender(encoder) encodes the image into bitstream $Y = f_0(X)$
3. Y is transmitted from sender to receiver

Encode and Decode procedure

1. The sender(encoder) sample image from source
2. The sender(encoder) encodes the image into bitstream
3. *Y* is transmitted from sender to receiver
4. Decoder inverses receiving *Y* using an **unconditional generative model** with idempotence constraint

sampling;
approximate the source

$$\min ||f_0(\hat{X}) - Y||^2, \text{ s.t. } \hat{X} \sim q_X$$

Encode and Decode procedure

1. The sender sample image from source
2. The sender encodes the image into bitstream
3. *Y* is transmitted from sender to receiver
4. Decoder inverses receiving *Y* using an **unconditional generative model** with idempotence constraint

sampling;
approximate the source

$$\min ||f_0(\hat{X}) - Y||^2, \text{ s.t. } \hat{X} \sim q_X$$

- Most generative inversion use the gradient:

$$\nabla_{\hat{X}} ||f_0(\hat{X}) - \boxed{Y}||^2$$

Not differentiable

Encode and Decode procedure

1. The sender sample image from source
2. The sender encodes the image into bitstream
3. *Y* is transmitted from sender to receiver
4. Decoder inverses receiving *Y* using an **unconditional generative model** with idempotence constraint

sampling;
approximate the source

$$\min ||f_0(\hat{X}) - Y||^2, \text{ s.t. } \hat{X} \sim q_X$$

- Most generative inversion use the gradient:

$$\nabla_{\hat{X}} ||f_0(\hat{X}) - \boxed{Y}||^2$$

Not differentiable

** For practical implementation,

- x-domain constraint => $\min ||g_0(f_0(\hat{X})) - g_0(Y)||^2, \text{ s.t. } \hat{X} \sim q_X$

Unconditional generative model procedure

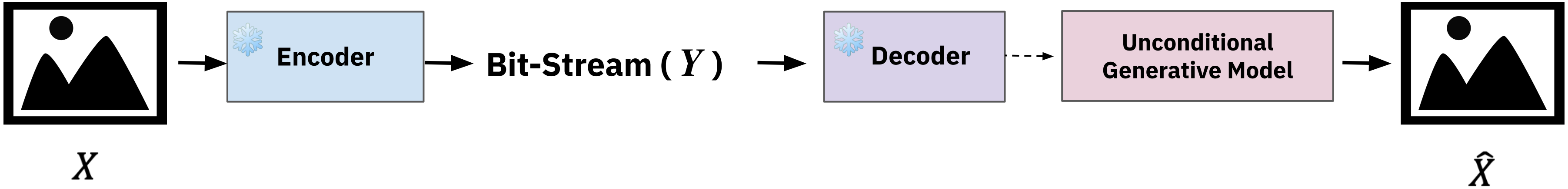
- StyleGAN2 + {PULSE, ILO}
- DDPM + {MCG, DPS}

Overall Architecture

Pre-trained MSE optimized (encoder, decoder) **+ Unconditional Generative Model**

- ELIC
- Hyper

- DDPM + DPS



Experiment Setup

- Evaluation
    - 1000 images of FFHQ and ImageNet validation split
    - metrics:
        - MSE
        - BD-metrics (Bjontegaard, 2001)
            - BD-FID, BD-PSNR

- Train (only unconditional generative model)
    - FFHQ (the remaining unused data for testing)
    - ImageNet (training split)

Experiment Setup

- Baselines
  - Perceptual Optimized ⇒ use conditional generative model
    - HiFiC
    - Po-ELIC
    - CDC
    - ILLM
  - MSE Optimized
    - Hyper
    - ELIC
  - (Handcraft)
    - VTM
    - BPG

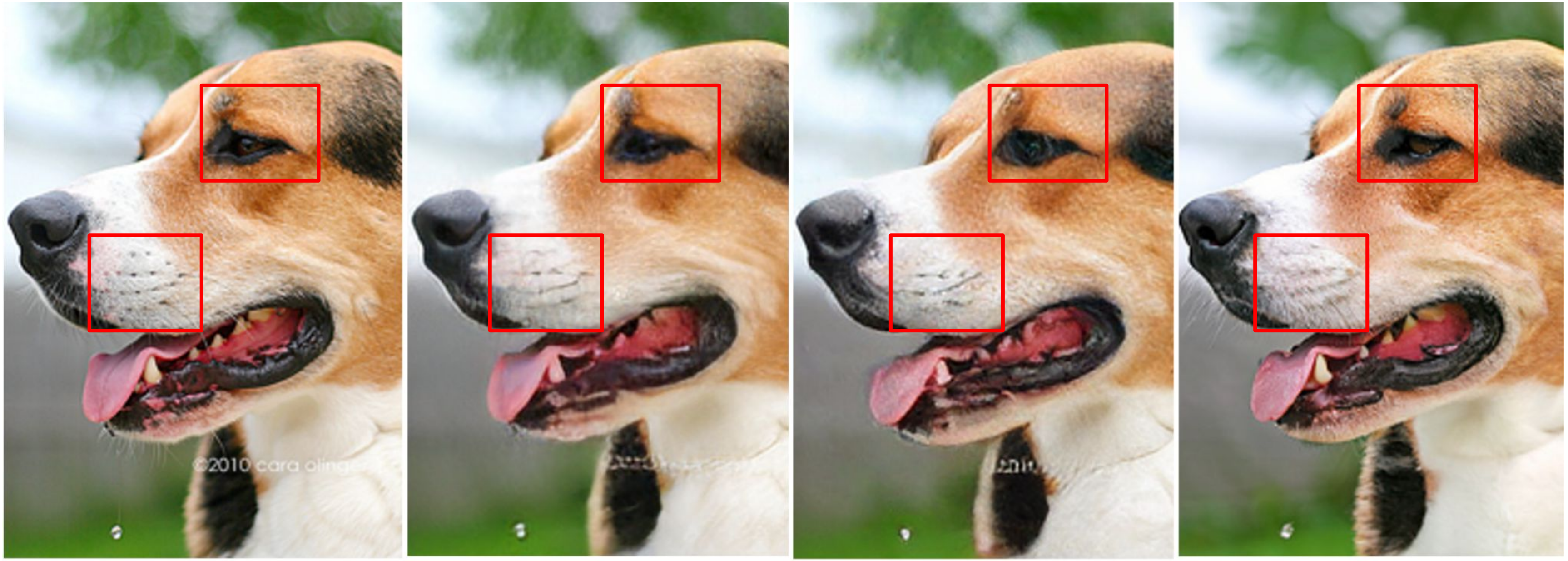
Result : FID, PSNR

- Proposed method outperforms HiFiC and ILLM in terms of FID.
    - PSNR result could be explained by perception-distortion trade-off.

Method	FFHQ		ImageNet		COCO		CLIC	
	BD-FID ↓	BD-PSNR ↑	BD-FID ↓	BD-PSNR ↑	BD-FID ↓	BD-PSNR ↑	BD-FID ↓	BD-PSNR ↑
MSE Baselines								
Hyper	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
ELIC	-9.740	1.736	-10.50	1.434	-8.070	1.535	-10.23	1.660
BPG	-4.830	-0.8491	-8.830	-0.3562	-4.770	-0.3557	-4.460	-0.4860
VTM	-14.22	0.7495	-13.11	0.9018	-11.22	0.9724	-12.21	1.037
Conditional Generative Model-based								
HiFiC	-48.35	-2.036	-44.52	-1.418	-44.88	-1.276	-36.16	-1.621
HiFiC*	-51.85	-1.920	-47.18	-1.121	-	-	-	-
Po-ELIC	-50.77	0.1599	-48.84	0.1202	-50.81	0.2040	-42.96	0.3305
CDC	-43.80	-8.014	-41.75	-6.416	-45.35	-6.512	-38.31	-7.043
ILLM	-50.58	-1.234	-48.22	-0.4802	-50.67	-0.5468	-42.95	-0.5956
ILLM*	-52.32	-1.415	-47.99	-0.7513	-	-	-	-
Unconditional Generative Model-based								
Proposed (Hyper)	<u>-54.14</u>	-2.225	<u>-52.12</u>	-2.648	<u>-56.70</u>	-2.496	<u>-44.52</u>	-2.920
Proposed (ELIC)	**-54.89**	-0.9855	**-55.18**	-1.492	**-58.45**	-1.370	**-46.52**	-1.635

Table 2: Results on FFHQ, ImageNet, COCO and CLIC. *: re-trained on corresponding dataset. **Bold**: lowest FID. <u>Underline</u>: second lowest FID.

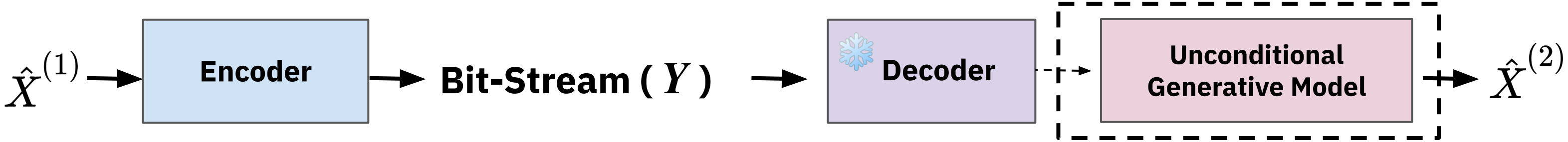
Qualitative Result



Source (ImageNet) | HiFiC 0.17 bpp | ILLM 0.11 bpp | Proposed 0.11 bpp

Re-compression Experiment

- Evaluate idempotence by MSE between first time compression and re-compression.



Result : MSE of Re-compression

- Proposed method re-compression MSE is **smaller** than the existed MSE optimized methods.

  ⇒ improve idempotence !

	Re-compression metrics	
	MSE ↓	PSNR (dB) ↑
Hyper	6.321	40.42
Hyper w/ Proposed	2.850	44.84
ELIC	11.80	37.60
ELIC w/ Proposed	7.367	40.93

Limitation: Diversity of reconstruction

⚠️ **Note.** Differ a lot in detail but the authors say all have good visual quality

Original (COCO) Alternative reconstructions

Q&A

Appendix: Result of KID, LPIPS

Method	FFHQ		ImageNet		COCO		CLIC	
	BD-logKID ↓	BD-LPIPS ↓	BD-logKID ↓	BD-LPIPS ↓	BD-logKID ↓	BD-LPIPS ↓	BD-logKID ↓	BD-LPIPS ↓
MSE Baselines								
Hyper	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
ELIC	-0.232	-0.040	-0.348	-0.058	-0.236	-0.062	-0.406	-0.059
BPG	0.1506	-0.010	0.027	-0.010	0.126	-0.012	0.039	-0.008
VTM	-0.232	-0.031	-0.298	-0.048	-0.216	-0.050	-2.049	-0.048
Conditional Generative Model-based								
HiFiC	-3.132	-0.108	-2.274	-0.172	-2.049	-0.172	-1.925	-0.148
HiFiC*	-4.261	-0.110	-2.780	-0.173	-	-	-	-
Po-ELIC	-3.504	-0.104	-2.877	-0.167	-2.671	-0.168	-2.609	-0.145
CDC	-2.072	-0.060	-1.968	-0.099	-1.978	-0.101	-2.122	-0.084
ILLM	-3.418	-0.109	-2.681	-0.181	-2.620	-0.180	-2.882	-0.155
ILLM*	-4.256	-0.106	-2.673	-0.178	-	-	-	-
Unconditional Generative Model-based								
Proposed (Hyper)	<u>-5.107</u>	-0.086	<u>-4.271</u>	-0.058	<u>-4.519</u>	-0.083	<u>-3.787</u>	-0.056
Proposed (ELIC)	**-5.471**	-0.099	**-5.694**	-0.106	**-5.360**	-0.113	**-4.046**	-0.079

Table 7: Results on FFHQ, ImageNet, COCO and CLIC. *: re-trained on corresponding dataset. **Bold**: lowest KID. <u>Underline</u>: second lowest KID.

Appendix: Qualitative Result