The Lossy Compression Theory

Jianping Lin

2020/04/08

The Lossy Compression Theory

- **Currently, the lossy compression theory can be divided into three classes according to different objectives**

Shannon's Rate-Distortion Theory

- There is a two-way tradeoff between rate and distortion (signal fidelity: MSE, MS-SSIM, any other full reference criterion) .
- The goal is to achieve the lowest possible distortion at a given rate.

[1] Shannon, C. E. Coding theorems for a discrete source with a fidelity criterion. IRE Nat. Conv. Rec, 4(142-163):1, 1959.

The Rate-Distortion-Perception Tradeoff

- It is a generalized rate-distortion theory which takes perceptual quality into account.
- It characterizes the three-way tradeoff between rate, distortion, and perception .

[2] Blau, Yochai, and Tomer Michaeli. "Rethinking lossy compression: The rate-distortion-perception tradeoff." ICML (2019).

The Classification-Distortion-Perception Tradeoff

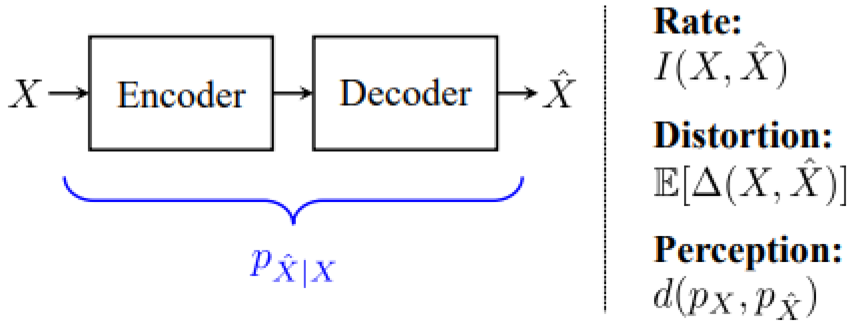
- It further takes the semantic quality of the restored signal into account (the utility of the signal for recognition purpose).
- There is a three-way tradeoff between distortion, perception, and classification, at any given rate.

[3] Liu, Dong, Haochen Zhang, and Zhiwei Xiong. "On The Classification-Distortion-Perception Tradeoff." Advances in Neural Information Processing Systems. 2019.

The Rate-Distortion-Perception Tradeoff

Background

Rate-Distortion Theory



- As shown in the above figure, for an inputted iid source $X \sim p_X$, the expected distortion between X and the decoded signal $\hat{X}$ is defined as:

$$\mathbb{E}[\Delta(X, \hat{X})],$$

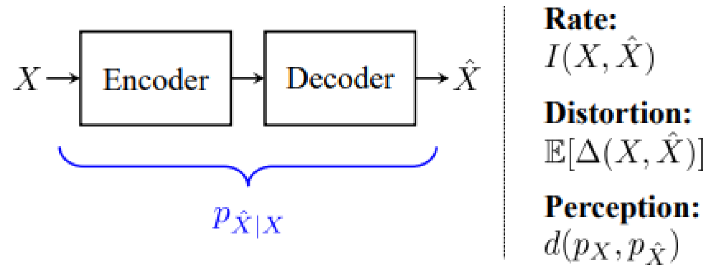
where the expectation is with respect to the joint distribution $p_{X,\hat{X}} = p_{\hat{X}|X} p_X$, and $\Delta : \mathcal{X} \times \hat{\mathcal{X}} \to \mathbb{R}^+$ is any full reference distortion measure(e.g., squared error, MS-SSIM, etc.)

[2] Blau, Yochai, and Tomer Michaeli. "Rethinking lossy compression: The rate-distortion-perception tradeoff." ICML (2019).

The Rate-Distortion-Perception Tradeoff

Background

Rate-Distortion Theory



- If the expected distortion is bounded by D , then the lowest achievable rate R is characterized by the rate-distortion function:

$$R(D) = \min_{p_{\hat{X}|X}} I(X, \hat{X}) \quad \text{s.t.} \quad \mathbb{E}[\Delta(X, \hat{X})] \leq D,$$

Where I denotes mutual information.

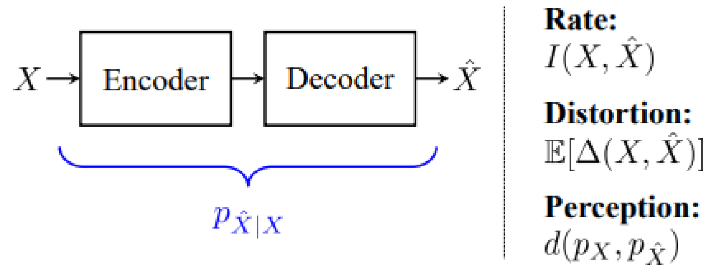
- Closed form expressions for $R(D)$ are known for only a few source distributions and under quite simple distortion measures (e.g. squared error or Hamming distance).

- Several general properties are known, including that it is always monotonically non-increasing and convex.

The Rate-Distortion-Perception Tradeoff

Background

Perceptual Quality



- The perceptual quality of an output sample $\hat{x}$ refers to the extent to which it is perceived by humans as a valid (natural) sample, regardless of its similarity to the input x. (no-reference, perceptual naturalness)
- Based on previous works, the authors define the perceptual quality index (lower is better) as:

$$d(p_X, p_{\hat{X}}),$$

where d(.,.) is some divergence between distributions (e.g. Kulback-Leibler, Wasserstein, etc.).

They assume that $d(p,q) \geq 0, d(p,q) = 0 \Leftrightarrow p = q.$

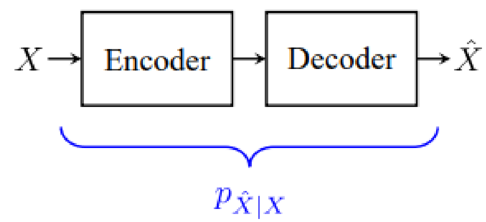
- Obviously, perceptual quality, as defined above, is very different from distortion. In particular, minimizing the perceptual quality index does not necessarily lead to low distortion, and minimizing distortion does not necessarily lead to good perceptual quality, neither.
- In fact, perception and distortion are fundamentally at odds with each other (for non-invertible degradations), in the sense that optimizing one always comes on the expense of the other.

The Rate-Distortion-Perception Tradeoff

The Rate-Distortion-Perception Tradeoff

Definition 1 *The (information) rate-distortion-perception function is defined as*

$$R(D,P) = \min_{p_{\hat{X}|X}} I(X,\hat{X})$$
$$\text{s.t. } \mathbb{E}[\Delta(X,\hat{X})] \le D, \;\; d(p_X, p_{\hat{X}}) \le P. \quad (4)$$



Rate: $I(X,\hat{X})$

Distortion: $\mathbb{E}[\Delta(X,\hat{X})]$

Perception: $d(p_X, p_{\hat{X}})$

- The closed form solutions for (4) are even harder to obtain. Therefore, the authors study the classical case of a binary source(Bernoulli source) to understand the properties of this function.

Bernoulli Source

- Consider the problem of encoding a binary source $X \sim Bern(p)$, where the decoder's output $\hat{X}$ is also constrained to be binary. Take the Hamming distance as the distortion measure $\Delta(.,.)$,and the total-variation (TV) distance $d_{TV}(.,.)$ as the perception index.
- When perception is not constrained (i.e., P = ∞), the solution is:

$$R(D,\infty) = \begin{cases} H_b(p) - H_b(D) & D \in [0,p) \\ 0 & D \in [p,\infty) \end{cases} \quad (5)$$

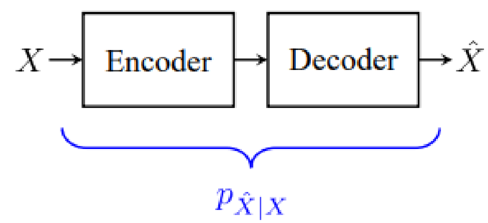
where $H_b(\alpha)$ is the entropy of a Bernoulli random variable with probability α

The Rate-Distortion-Perception Tradeoff

□ The Rate-Distortion-Perception Tradeoff

Definition 1 *The (information) rate-distortion-perception function is defined as*

$$R(D,P) = \min_{p_{\hat{X}|X}} I(X,\hat{X}) \quad \text{s.t. } \mathbb{E}[\Delta(X,\hat{X})] \leq D,\ d(p_X, p_{\hat{X}}) \leq P. \quad (4)$$



Rate: $I(X,\hat{X})$

Distortion: $\mathbb{E}[\Delta(X,\hat{X})]$

Perception: $d(p_X, p_{\hat{X}})$

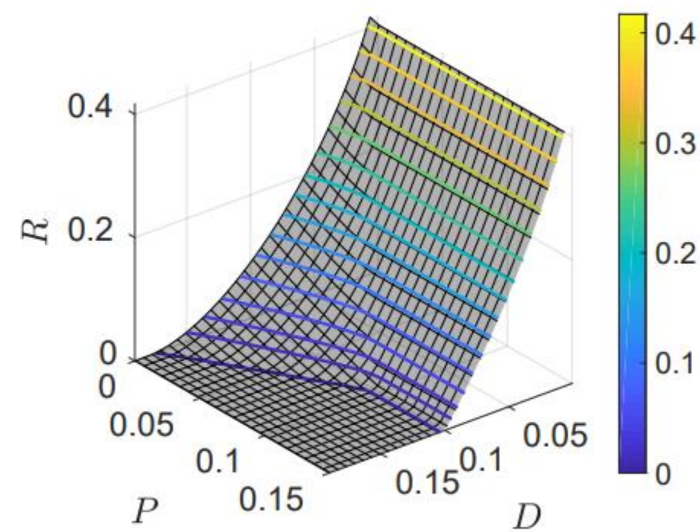
■ Bernoulli Source

- When perception is constrained by arbitrary P, if the perceptual quality constraint is sufficiently loose, the solution remains the same with (5). However, when $P \leq p$, the perception constraint in (4) becomes active and the solution is:

$$R(D,P) = \quad (6)$$

$$\begin{cases} H_b(p) - H_b(D) & D \in \mathcal{S}_1 \\ 2H_b(p) + H_b(p-P) - H_t(\frac{D-P}{2}, p) - H_t(\frac{D+P}{2}, q) & D \in \mathcal{S}_2 \\ 0 & D \in \mathcal{S}_3 \end{cases}$$

where $q = 1 - p$ and $H_t(\alpha, \beta)$ denotes the entropy of a ternary random variable with probabilities $\alpha, \beta, 1-\alpha-\beta$. Here, $\mathcal{S}_1 = [0, D_1)$, $\mathcal{S}_2 = [D_1, D_2)$, and $\mathcal{S}_3 = [D_2, \infty)$, where $D_1 = \frac{P}{1-2(p-P)}$ and $D_2 = 2pq - (q-p)P$.

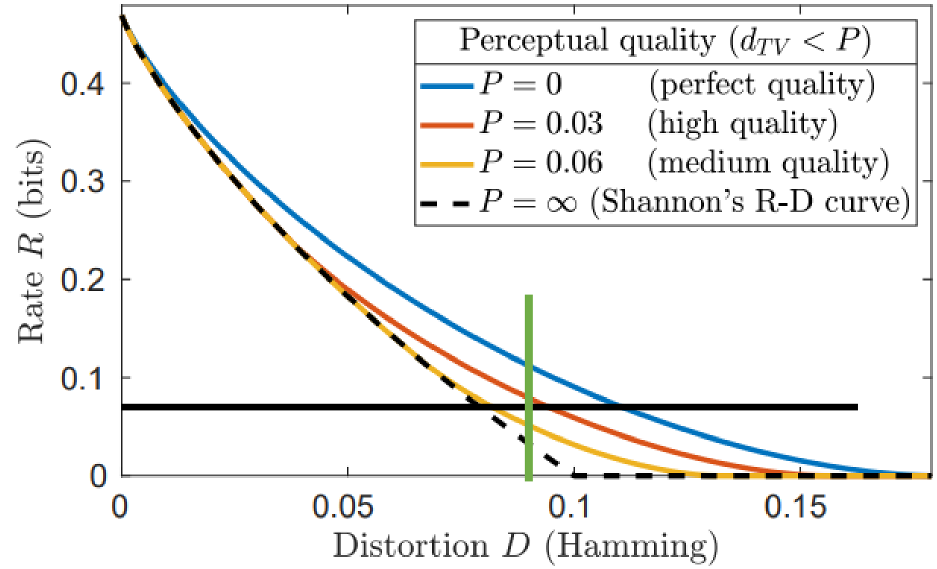


The Rate-Distortion-Perception Tradeoff

☐ The Rate-Distortion-Perception Tradeoff

■ Bernoulli Source

- This figure displays cross-sections of R(D,P) along rate-distortion planes.



- The lower the perceptual quality constraint (higher P), the lower the rate-distortion elevates (colored curves). When no constraint (P = ∞), it becomes the Shannon's R-D curve.

- At a given rate (black line), as the perceptual quality constraint becomes higher (lower P), the distortion D becomes larger .

- Similarly, for a fixed distortion D (green line), as the perceptual quality constraint becomes higher (lower P), the rate R becomes larger .

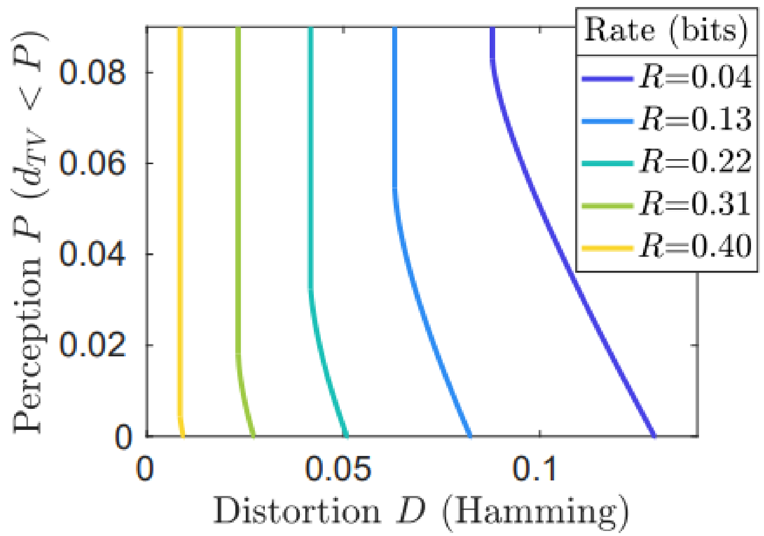
The Rate-Distortion-Perception Tradeoff

The Rate-Distortion-Perception Tradeoff

Bernoulli Source

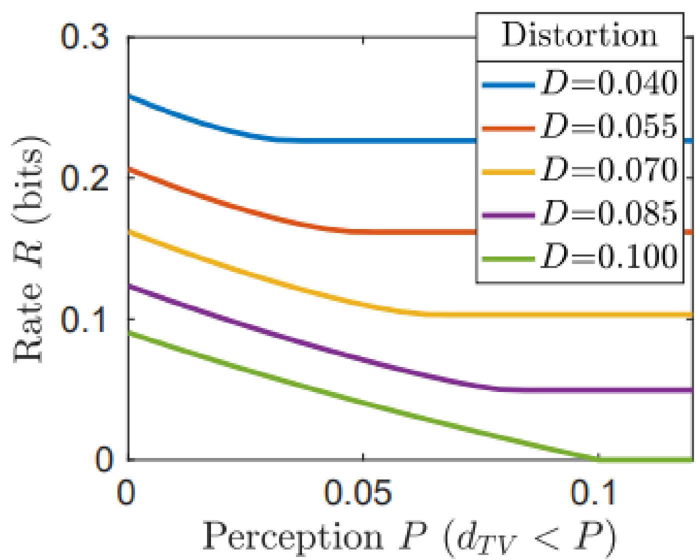
- This figure displays cross-sections of R(D,P) along perception-distortion planes.

- The tradeoff between distortion and perceptual quality becomes stronger at low bit-rates



- This figure displays cross-sections of R(D,P) along rate-perception planes.

- At every constant distortion level, the perceptual quality can be improved by increasing the rate.

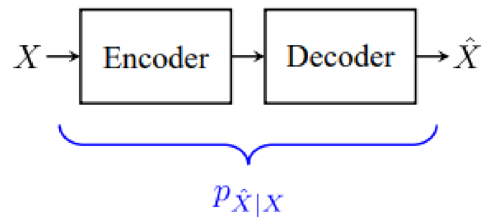


The Rate-Distortion-Perception Tradeoff

The Rate-Distortion-Perception Tradeoff

Definition 1 *The (information) rate-distortion-perception function is defined as*

$$R(D,P) = \min_{p_{\hat{X}|X}} I(X,\hat{X})$$
$$\text{s.t. } \mathbb{E}[\Delta(X,\hat{X})] \leq D,\ d(p_X, p_{\hat{X}}) \leq P. \quad (4)$$



Rate:
$I(X,\hat{X})$

Distortion:
$\mathbb{E}[\Delta(X,\hat{X})]$

Perception:
$d(p_X, p_{\hat{X}})$

Theoretical Properties

- For general source distributions, it is usually impossible to solve (4) analytically. However, it turns out that the behavior for a Bernoulli source is quite typical. They next prove several general properties of the function (4), which hold under rather mild assumptions. Specifically, they assume:

A1 The divergence $d(\cdot,\cdot)$ in (4) is convex in its second argument. That is, for any $\lambda \in [0,1]$ and for any three distributions p_0, q_1, q_2,

$$d(p_0, \lambda q_1 + (1-\lambda) q_2) \leq \lambda d(p_0, q_1) + (1-\lambda) d(p_0, q_2). \quad (7)$$

A2 The function $k(z) = \mathbb{E}_{X\sim p_X}[\Delta(X,z)]$ is not constant[5] over the entire support of p_X.

- Any f-divergence (e.g. KL, TV, Hellinger) satisfies assumption A1

- A valid signal is any $x : p_X(x) > 0.$ for assumption A2

- Using these assumptions, they are able to qualitatively characterize the general shape of the function R(D, P).

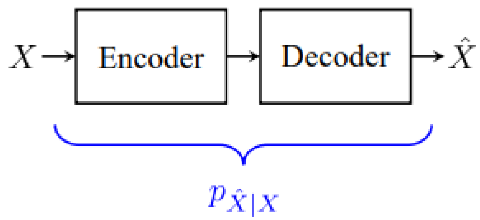
The Rate-Distortion-Perception Tradeoff

The Rate-Distortion-Perception Tradeoff

Definition 1 *The (information) rate-distortion-perception function is defined as*

$$R(D,P) = \min_{p_{\hat{X}|X}} \; I(X,\hat{X})$$

$$\text{s.t. } \mathbb{E}[\Delta(X,\hat{X})] \le D, \;\; d(p_X, p_{\hat{X}}) \le P. \quad (4)$$



Rate: $I(X,\hat{X})$

Distortion: $\mathbb{E}[\Delta(X,\hat{X})]$

Perception: $d(p_X, p_{\hat{X}})$

Theoretical Properties

Theorem 1 *The rate-distortion-perception function* (4)*:*

1. *is monotonically non-increasing in D and P;*
2. *is convex if **A1** holds;*
3. *satisfies $R(\cdot,0) \neq R(\cdot,\infty)$ if **A2** holds.*

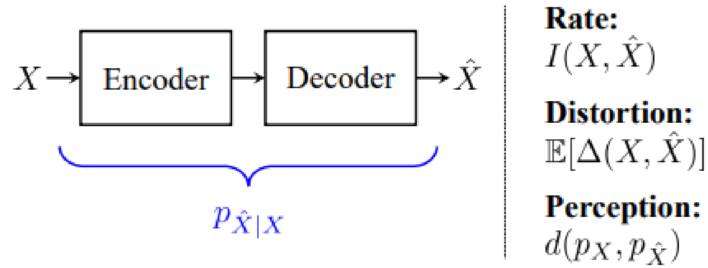
- Properties 1 and 3 indicate that there exists some D0 for which R(D_0, 0) > R(D_0, ∞), showing that the rate-distortion curve necessarily elevates when constraining for perfect perceptual quality, whatever the distortion criterion is.

The Rate-Distortion-Perception Tradeoff

☐ The Rate-Distortion-Perception Tradeoff

Definition 1 *The (information) rate-distortion-perception function is defined as*

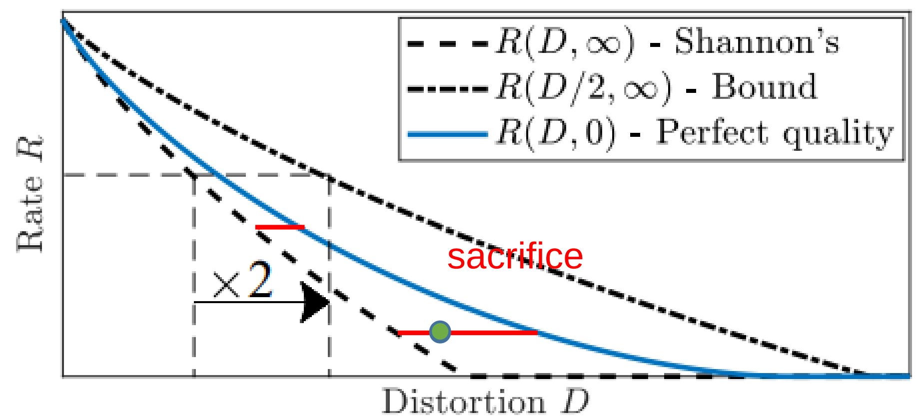
$$R(D,P) = \min_{p_{\hat{X}|X}} I(X,\hat{X})$$
$$\text{s.t. } \mathbb{E}[\Delta(X,\hat{X})] \le D,\ d(p_X,p_{\hat{X}}) \le P. \quad (4)$$



■ Theoretical Properties

Theorem 2 *When using the squared-error distortion, the function $R(\cdot,0)$ (rate-distortion at perfect perceptual quality) is bounded by*

$$R(D,0) \le R(\tfrac{1}{2}D,\infty). \quad (8)$$



- ➢ Theorem 2 shows that it is possible to attain perfect perceptual quality without increasing the rate, by sacrificing no more than a 2-fold increase in the MSE (3dB decrease in PSNR). The lower rate, the more sacrifice in MSE.
- ➢ This bound is generally not tight. Thus, in some settings, perfect perceptual quality can be obtained with an even smaller increase in 2x distortion. In practice, if we change the optimization of our model from MSE to perceptual quality, maybe we can obtain the operating point from the green circle to the red circle without increasing the rate but attaining a much better perceptual quality.

The Rate-Distortion-Perception Tradeoff

Experimental Illustration

- They develop an auto-encoder based lossy image compression on a toy MNIST example. The rate is controlled by the dimension dim of the encoder's output $f(X)$, and the number of levels L used for quantizing each of its entries, such that $R \leq dim \times \log_2(L)$

- For any fixed rate, they train the encoder-decoder to minimize a loss comprising a weighted combination of the expected distortion and the perception index,

$$\mathbb{E}[\Delta(X, g(\hat{f}(X)))] + \lambda d_{\mathrm{W}}(p_X, p_{\hat{X}}).$$

where $d_W(.,.)$ is the Wasserstein distance. The perceptual quality term can be optimized with the framework of generative adversarial networks (GAN) by introducing an additional (discriminator) DNN $h : \mathcal{X} \to \mathbb{R}$ and minimizing

$$\mathbb{E}[\Delta(X, g(\hat{f}(X)))] + \lambda \max_{h \in \mathcal{F}} \Big(\mathbb{E}[h(X)] - \mathbb{E}[h(g(\hat{f}(X)))] \Big)$$

- To achieve good perceptual quality, especially at low rates, it is essential that the decoder be stochastic. They add a noise vector to the encoder's output $\hat{f}(x) + n$.

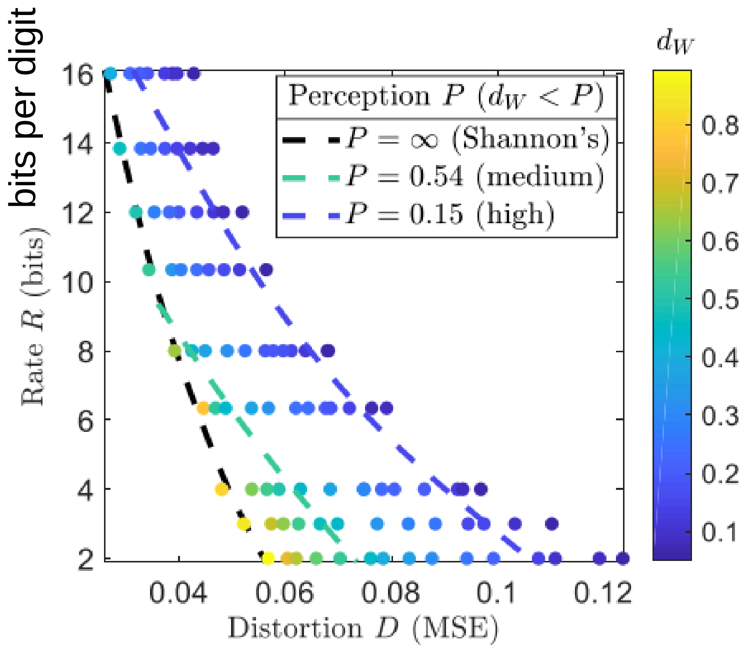
Squared-Error Distortion

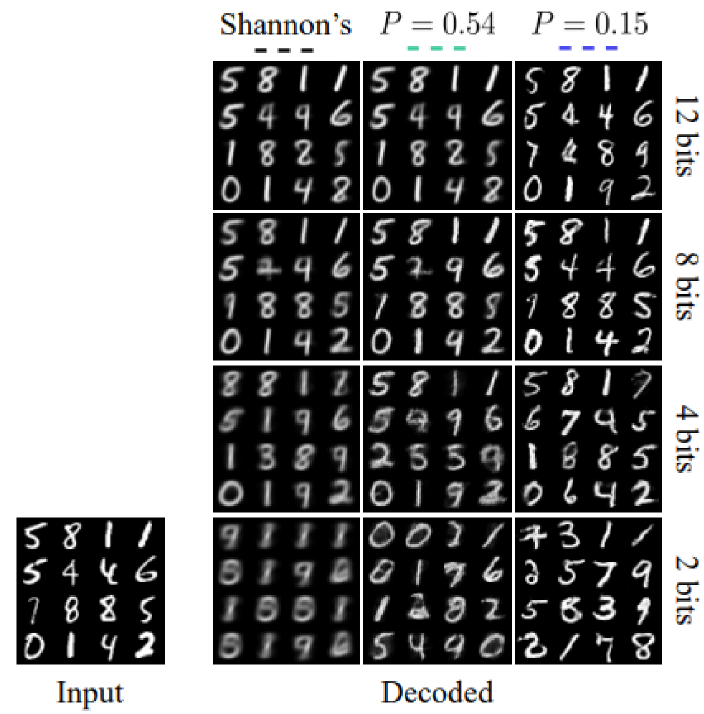
- They begin by experimenting with the squared-error distortion $\Delta(x, \hat{x}) = \|x - \hat{x}\|^2$.

- They train 98 encoder-decoder pairs on the MNIST handwritten digit dataset, while varying dim and L to control the rate R, and the tuning coefficient λ to achieve different balances between distortion and perceptual quality.

The Rate-Distortion-Perception Tradeoff

☐ Experimental Illustration





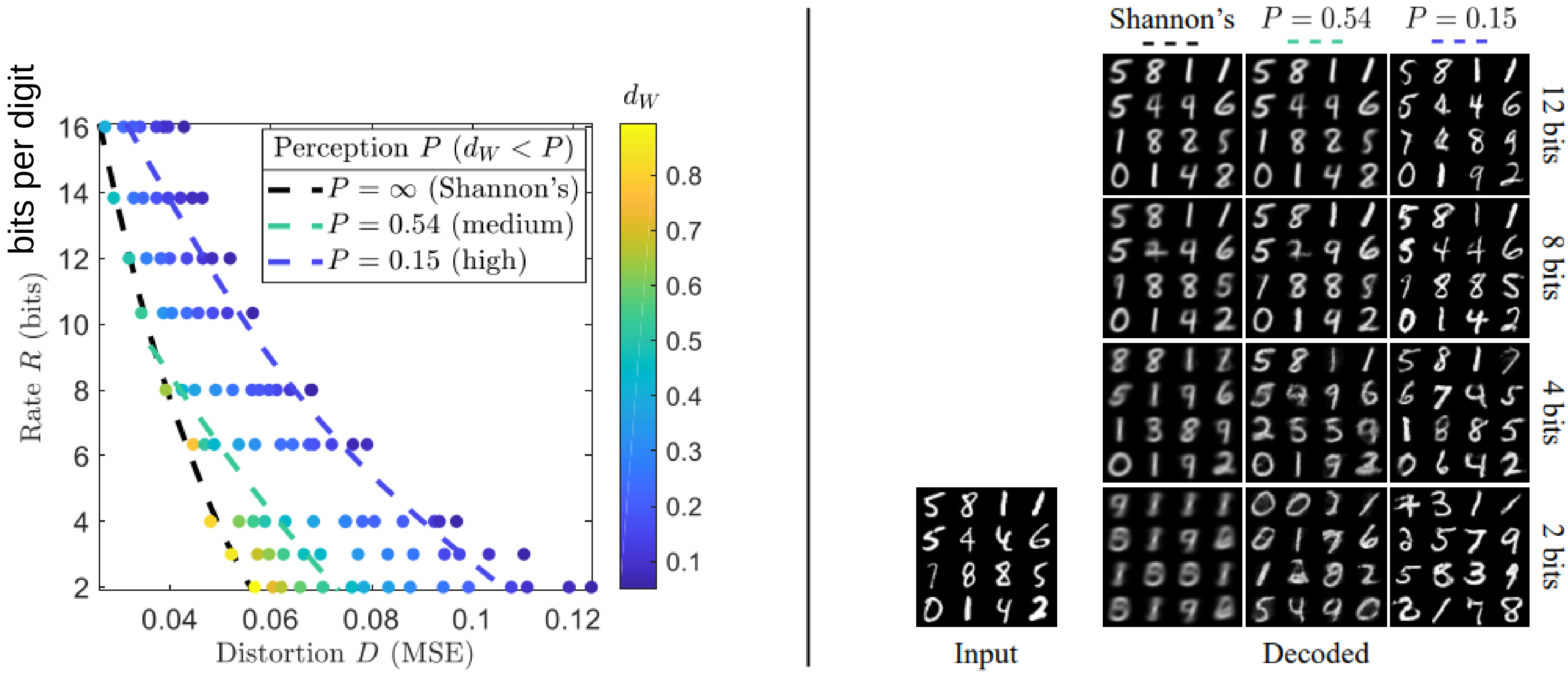
➢ They plot two smoothing spline calculated over the set of points which satisfy the constraint $d_W(p_X, p_{\hat{X}}) \leq P$ and have the minimal distortion among all points with the same rate. The Shannon's rate-distortion function is obtained with λ = 0

$$^{7}\mathcal{L}_{\text{dis}} \quad = \quad \frac{2}{N}\left(\sum_{i=1}^{N/2} h(x_i) - \sum_{i=N/2+1}^{N} h(g(\hat{f}(x_i)))\right),$$

where $\{x_i\}_{i=1}^{N}$ are the test samples.

The Rate-Distortion-Perception Tradeoff

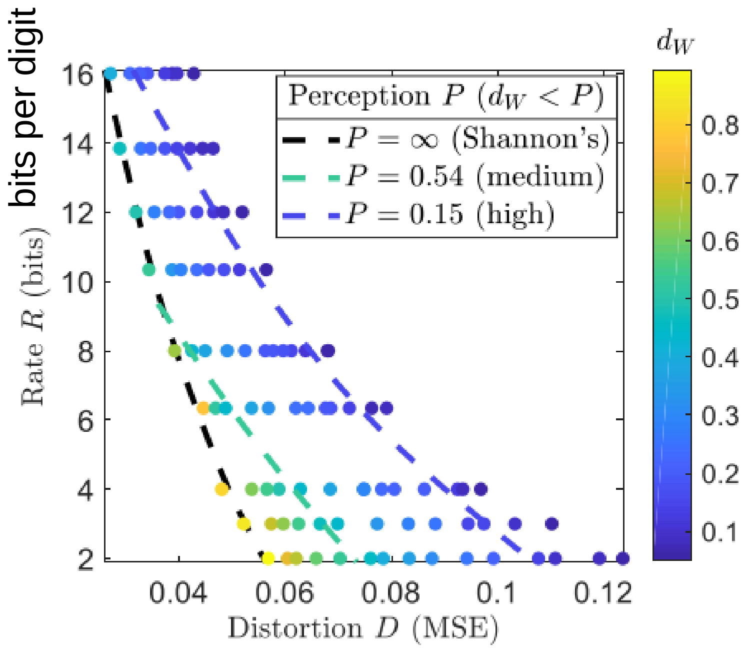
☐ Experimental Illustration

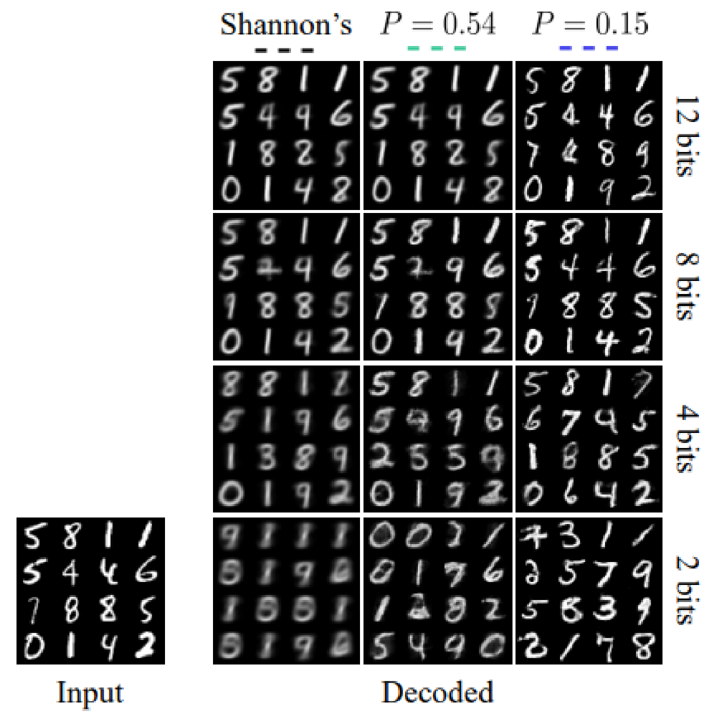


➢ It demonstrates once again that we can improve the perceptual quality w.r.t. that obtained on Shannon's rate-distortion curve, yet this must come at the cost of a higher rate and/or distortion. Notice that the perception index is not constant along Shannon's function; it increases (worse quality) towards lower bit-rates.

The Rate-Distortion-Perception Tradeoff

☐ Experimental Illustration

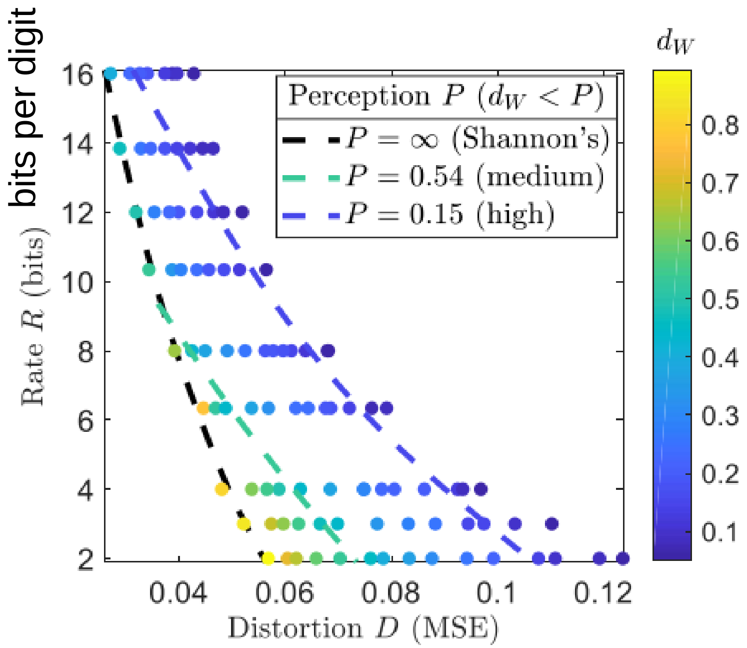


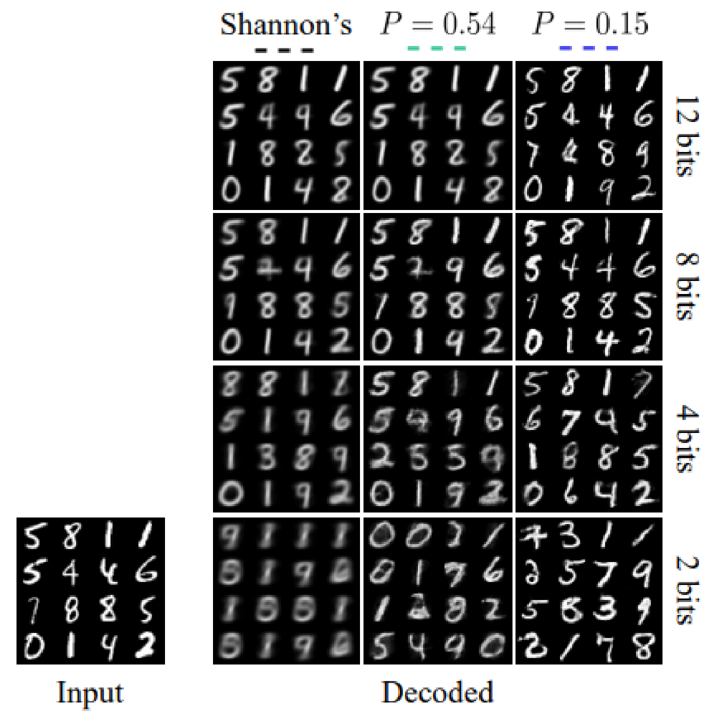


➢ It demonstrates once again that we can improve the perceptual quality w.r.t. that obtained on Shannon's rate-distortion curve, yet this must come at the cost of a higher rate and/or distortion. Notice that the perception index is not constant along Shannon's function; it increases (worse quality) towards lower bit-rates.

The Rate-Distortion-Perception Tradeoff

☐ Experimental Illustration





➢ As the rate decreases, the perceptual quality of the reconstructions along Shannon's function degrades. However, this is avoided when constraining the perceptual quality, which results in visually pleasing reconstructions even at extremely low bit-rates. This increased perceptual quality does not imply increased accuracy, as at low bit rates (e.g., 2 bits), most reconstructions fail to preserve even the identity of the digit.

The Rate-Distortion-Perception Tradeoff

Experimental Illustration

Advanced Distortion Measures

- Apart from the MSE distortion, there are other distortion criteria more consistent with human vision (e.g., MS-SSIM, deep-feature based distortion, etc.)
- Yet, the perceptual quality along Shannon's rate-distortion function is not perfect for nearly any distortion measure (see property 3 in Theorem 1). This implies that perfect perceptual quality cannot be achieved by merely switching to more advanced distortion criteria, but rather requires directly optimizing the perception index (e.g. using GAN-based schemes). The strength of the tradeoff can certainly decrease for distortion criteria which capture more semantic similarities.

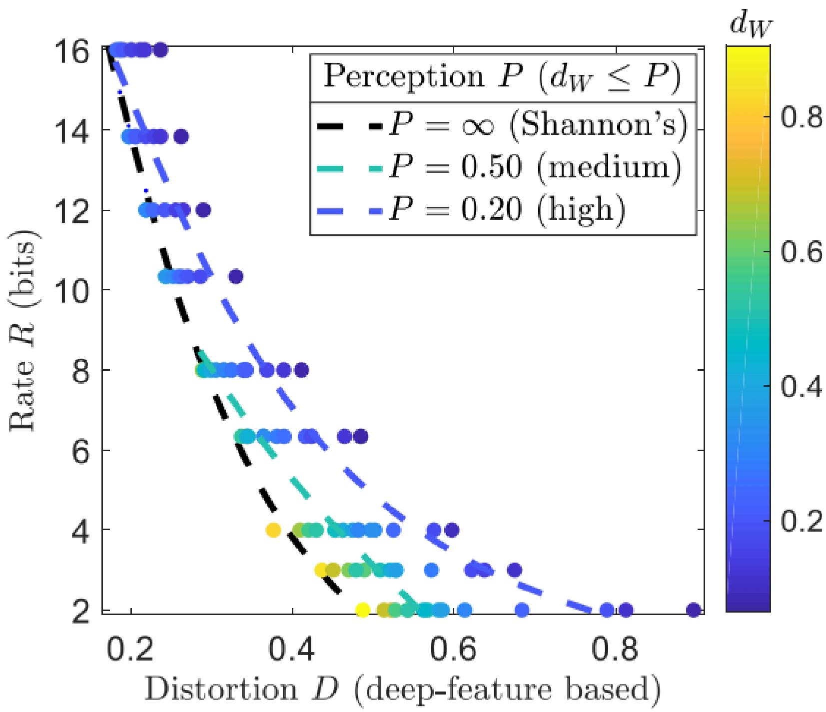
- To demonstrate this, they try the deep-feature based distortion

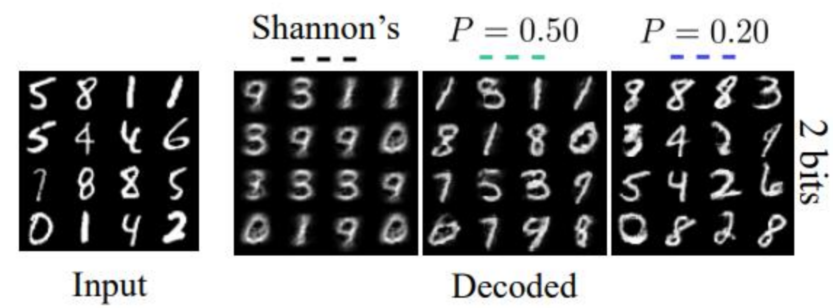
$$\Delta(x, \hat{x}) = \|x - \hat{x}\|^2 + \alpha \|\Psi(x) - \Psi(\hat{x})\|^2,$$

Here, they take the second conv-layer output of a 4-layer DNN, which they pre-trained to achieve over 99% classification accuracy on the MNIST test set

The Rate-Distortion-Perception Tradeoff

Experimental Illustration





- As can be seen, here too the rate-distortion curves elevate when constraining the perceptual quality, demonstrating that the use of advanced distortion measures does not eliminate the tradeoff. From the decoded outputs, however, it is evident that the tradeoff here is somewhat weaker, as minimizing distortion alone (Shannon's) appears a bit more visually pleasing.

The Rate-Distortion-Perception Tradeoff

Conclusion

- There is a triple tradeoff between rate, distortion and perception. Any attempt to keep the statistics of decoded signals similar to that of source signals, will result in a higher distortion or rate.

- Comparing methods based on their rate-distortion curves alone may be misleading, especially at low rate. A more informative evaluation must also include some (no-reference) perceptual quality measure.

- It is a very useful theory that can help us to develop the lossy compression method. For example, if we want to obtain a more visually pleasing result at the same rate, we can increase the weight of perception quality constraint, but it will increase the distortion accordingly. More flexibly, we can design a region-dependent image compression method. Specifically, we can first generate an region-dependent perception weight map at the encoder side and then transmit it to the decoder. Therefore, the decoder can reconstruct a visually pleasing result with variable-rate regions (i.e., the texture regions like grass/tree/water… can be visually pleasing with very low rate at a higher perceptual quality constraint (high distortion), and other structure regions like people/car/building… can be more accurate (lower distortion/ higher fidelity) with much higher rate at no/lower perceptual quality constraint).


Thank You!